

وزارة التعليم العالي والبحث العلمي
جامعة تكريت - كلية العلوم

مبادئ
علم الاحصاء التطبيقي
لغير الاختصاص

تاليف
غازي عطية زراك

رقم الايداع في دار الكتب والوثائق ببغداد

(1051) لسنة 2015

المؤلف : غازي عطية زراك

الطبعة الاولى

2015

مبادئ
علم الاحصاء التطبيقي
لغير الاختصاص

تاليف

المدرس

غازي عطية زراك

جامعة تكريت–كلية العلوم

بسم الله الرحمن الرحيم

وَأحاطَ بِما لَدَيّهم وَأحصىُ كُلَّ شَىءً عَدداُ (28)

"صدق الله العظيم"

(سورة الجن)

الاهداء

الى كل من يحمل هم الوطن ...

الذي ينتظر منهم التقدم والرقي ...

طلبة الحاضر والمستقبل ...

المحتويات

المقدمة

الحمد لله رب العالمين , والصلاة والسلام على خاتم النبيين وسيد المرسلين , نبينا محمد بن عبد الله الهادي البشير الذي بعثه الله رحمة للعالمين , وعلى آله واصحابه اجمعين , ومن اهتدى بهديه الى يوم الدين , وبعد .

يتعامل الناس عموماً في حياتهم اليومية مع المفاهيم وحتى بعض المفردات الإحصائية وبالأخص ما يتعلق منها بالإحتمالات وبعض المقاييس الوصفية مع رصد ما يطرأ عليها من تغيرات عبر فترات زمنية متعاقبة. وإذا كان هذا الأمر محسوساً بالنسبة لنا في الوقت الحاضر، فإن حقب التأريخ القديم وما بعده أفرزت أحداثاً تنم مجرياتها عن بدء استخدام الأساليب الإحصائية على أرض الواقع والتي يمكن إعتبارها أفكاراً تتناغم مع بعض أحدث الأساليب الإحصائية المعاصرة.

علم الاحصاء ليس غريبا علينا نحن المجتمع الاسلامي , حيث تمت الاستفادة من علم الاحصاء في بناء المجتمع والعمل على تنميته وتطوره وازدهاره , عبر تطور المجتمع الاسلامي منذ القدم ولغاية الوقت الحاضر , اذا ان اجراء الاحصائيات واتباع الاساليب والوسائل الاحصائية في جمع المعلومات عن الظواهر والمشكلات المختلفة تهدف الى اتخاذ القرارات ومعالجة المشاكل بوقت مبكر منذ بداية بناء المجتمع الاسلامي وبداية تطوره ونهوضه .

ومع تطور علم الرياضيات في القرن الثامن عشر وظهر بعض النظريات العلمية الهامة مثل نظرية الاحتمال التي كان لها الدور الكبير في تطور علم الاحصاء واكتسابه اهمية كبرى بحيث اصبح علما مستقلا وانتشر استخدامه وبدأ الاهتمام به من قبل العلماء في تطبيق النظريات والطرق والاساليب الاحصائية في الكثير من فروع العلم الحديث باعتباره الطريقة الصحيحة والاسلوب الامثل في البحث العلمي .

يعتبر علم الاحصاء في الوقت الحاضر واحدا من اهم العلوم الحديثة التي تلعب دورا حيويا في كثير من العلوم والدراسات المختلفة . كما يعتبر علم الاحصاء التطبيقي من اقدم العلوم حيث ظهر مع حاجة الانسان الاولى للتعامل مع القيم والاعداد لتسيير امور الحياة اليومية منذ ان بدأ الانسان يتعامل مع الالة والعلوم المختلفة في تسهيل مهمة ترتيب امور حياته اليومية . ومع التطور الكبير والمتسارع في العلوم كافة في اواخر القرن العشرين , تطور علم الاحصاء ليستفيد من تقنيات الحاسوب بشكل يجعله العلم الاكثر تداخلاً مع العلوم الاخرى المختلفة , حيث اصبح يستخدم علم الاحصاء في العلوم التجارية وعلوم الطب والهندسة والعلوم الصرفة والعلوم الانسانية بمختلف تخصصاتها دون استثناء , كما ساهم تطور عصر المعلومات والانفتاح العالمي الحديث في ابراز اهمية تفعيل عملية التعامل مع البيانات الاحصائية باسلوب يضمن السيطرة عليها وقراءتها ومعالجتها. هذه

كان لها الاثر الكبير على تطور علم الاحصاء , فضلا عن ذلك فقد اتجهت الكثير من العلوم الاخرى خاصة العلوم التطبيقية منها الى استخدام علم الاحصاء من خلال حصر البيانات وتصنيفها وتبويبها والتعامل معها احصائيا بغية الوصول الى فهم افضل ومعالجات موضوعية لغرض الحصول على نتائج موثوقة .

ان علم الاحصاء في اللغة يعني (العد الشامل) حيث انه يتعامل مع الاعداد او البيانات الكمية , فاذا اردنا جمع بيانات عن ظاهرة ما مثل دراسة واقع مدرسة معينة او جامعة او مستشفى , فان ذلك يتم باحدى الصورتين اما جمع بيانات بصورة كمية او عددية او جمع بيانات بصورة وصفية . علم الاحصاء يتعامل مع البيانات الكمية او العددية ويمكنه ايضا التعامل مع البيانات الوصفية , حيث ان علم الاحصاء يتعامل مع الظواهر ايا كان نوعها تعاملا كميا او وصفيا ايضاً , ذلك ان الارقام لا بد ان يكون لها مدلولات , فالتعامل الوصفي يترتب عليه التعامل الكمي والعكس بالعكس في كثير من الدراسات .

تم الحرص في هذا الكتاب على عدم التعرض الى المفاهيم الدقيقة للنظريات الاحصائية والتي تحتاج الى قدر كبير من التحليل الرياضي , كما تم الحرص على عرض المواضيع والنظريات الاحصائية بطريقة مبسطة تعتمد على الايضاح دون التعرض الى النظريات الاساسية واشتقاقاتها مع تقديم عدد كبير من الامثلة العملية التي تساعد على سهولة الفهم والاستيعاب للمواضيع الاحصائية . هذا الكتاب يتناول مبادئ طرق تحليل البيانات الاحصائية بطريقتين هما الاحصاء الوصفي والذي تم التطرق اليه في الباب الاول والذي يشتمل على شرح المفاهيم الاحصائية والقوانين الاحصائية واتي تضم مقاييس النزعة المركزية ومقاييس التشتت . اما الباب الثاني فيشتمل على الاحصاء الاستدلالي والذي يضم شرح مبادئ نظرية الاحتمال والتوزيعات الاحتمالية وتقديمها بطريقة مبسطة لا تحتاج الى رياضيات متقدمة .

الباب الثالث خصص الى شرح تفصيلي عن معامل الارتباط ومخطط الانتشار مع التطرق بصورة مبسطة الى القوانين التي تحكم هذه المعاملات وكيفية احتسابها , اما الباب الرابع فانه يختص بشرح نظرية المعاينة او نظرية النمذجة , مع تعاريف وافية الى مدلولاتها ومفردات هذه النظرية المهمة .

المؤلف

غازي عطية زراك

الباب الاول

الاحصاء الوصفي

الفصل الاول

مفهوم الاحصــــاء والرموز الاحصائية

يعرف الاحصاء بانه ذلك العلم الذي يعمل على استخدام الاسلوب العلمي في طرق جمع البيانات او النتائج , ثم تصنيفها وتبويبها وعرضها وتحليلها بهدف الحصول منها على استنتاجات وقرارت مناسبة, أو بكلام آخر هو العلم الذي يبحث في طرق جمع البيانات الصحيحة والدقيقة عن ظاهرة ما ,ثم تلخيص, وصف وتحليل هذه البيانات بغية استخراج النتائج منها واتخاذ القرارات المناسبة. لقد اصبح علم الاحصاء في الوقت الحاضر اداة ووسيلة فعالة في البحث العلمي وتصنيف البيانات ومعالجتها في جميع العلوم.

يمكن تقسيم علم الاحصاء الى قسمين:

1. الاحصاء الوصفي Descriptive Statistics :

يشتمل الاحصاء الوصفي على جمع البيانات وعرضها ومعالجتها ووصفها بصورة قياسات رقمية وحسابات رياضية وبيانية, ثم تنظيمها وعرضها وحساب بعض المقاييس الاحصائية لها بدون اعطاء اي استنتاج حول الظاهرة الكلية المدروسة.

2. الاحصاء الاستدلالي Inference Statistics:

وتشتمل هذه الطريقة على عمل استنتاجات حول المصدر الذي جمعت منه البيانات او النماذج ويعتمد اعتمادا كبيرا على نظرية الاحتمال (Theory of Probability) , هذا الاستنتاج قد لا يكون مؤكد بصورة مطلقة وقد يكون خطأ. يختص الاحصاء الاستدلالي باستخلاص وتفسير النتائج واتخاذ القرارات.

تطبيقات الاحصاء في مجال الجيولوجيا اصبح اداة ووسيلة فعالة في الكشف عن الغموض الذي يرافق اعمال الاستكشاف والتقييم المعدني والذي اصبح يسمى حاليا بالرياضيات التطبيقية وذلك لاستخداماته الواسعة في معالجة مختلف انواع البيانات والمعلومات المستحصلة في مختلف مجالات العلوم الجيولوجية والمعرفة العلمية الاخرى , ويتم ذلك من خلال معالجات احصائية للبيانات والنتائج المستحصلة من العمل الحقلي او من التحاليل المختبرية المختلفة.

المصطلحات الاحصائية:

- المجتمع Population : هو عبارة عن مجموعة من المفردات او المشاهدات او الاشخاص والتي نرغب في دراسة وتحليل خصائصه, وهناك نوعين: مجتمع محدود أو نهائي ومجتمع غير محدود أو لا نهائي.
- العينة Sample: وهي عبارة عن مجموعة جزئية من المجتمع , وتعبر عنه اصدق تمثيل.
- البيانات Data: عبارة عن مجموعة من القيم او القياسات للمتغير الذي يرافق المفردات او عناصر المجتمع. قد تكون هذه البيانات على شكل ارقام او صفات او رموز.
- البيانات الاسمية : عبارة عن اسم او وصف لاي عنصر او مفردة في المجتمع.
- البيانات المنفصلة Discrete data: عبارة عن قيم تدل على صفة يمكن قياسها , وتاخذ جميع القيم الصحيحة والكسرية .

جمع البيانات الاحصائية:

إن جمع البيانات او النتائج او القراءات من ظاهرة معينة او مجموعة معينة تسمى بالمجتمع (Population) , مثل جمع نتائج قياس اطوال الطلبة في المرحلة الرابعة , او جمع نتائج قياس سمك طبقة رسوبية معينة , هذه النتائج او القراءات المستحصلة تسمى نماذج (Samples) وتعتبر جزء من الكل , هذا الجزء يجب ان يمثل الظاهرة المدروسة او الوسط الذي جمعت منه تلك النماذج, وبذلـك يمكن تعريف النموذج Sample بانه جزء من الكل يعطي او يعكس كامل خصائص ومميزات الكل. هذه النماذج تختلف من نموذج الى اخر , بغض النظر عن طبيعة هذه النماذج سواء كانت قياس لاطوال الطلبة, قياس وزن الطلبة, نسبة الامطار, سمك طبقة معينة أو قياس لتركيز الترسبات المعدنية, الى غير ذلك...... كل من هذه النماذج يسمى متغير (Variable) يختلف من نموذج الى اخر, وعادة يرمز لهذا المتغير بالرمز (Y) أو (X) .... الخ , كل نموذج وكل متغير ياخذ رقم محدد مثل النتغير الاول يرمز له بالرمز ($X1$) والمتغير الثاني ($X2$) .... وهكذا. مثال على ذلك اذا كان لدينا سبعة ابار لبابية محفورة لدراسة وتحديد سمك طبقة التربة العليا في منطقة معينة , نطلق على هذه المتغيرات السبعة التي تمثل السمك هي : $X1$, $X2$, $X3$, $X4$, $X5$, $X6$ وهذا , لكل متغير من هذه المتغيرات قيمة حقيقية خاصة به تمثل سمك الطبقة العليا. لنفترض ان :

$X1 = 20m$

$X2 = 25m$

$X3 = 10m$

$X4 = 15m$

X5 = 16m

X6 = 22m

X7 = 18m

هذه المتغيرات كل منهما يسمى نموذج وان طبقة التربة العليا هي الظاهرة المدروسة او التي نهتم بدراستها التي تسمى (Population) , هذا المفهوم ينطبق على اي متغير واي مجموعة شاملة تخضع للدراسة.

طبيعة البيانات الاحصائية:

تقسم المتغيرات Variables او النماذج الى قسمين:

1. متغيرات وصفية Qualitative Variables :

وهي الظواهر او الصفات التي لا يمكن قياسها مباشرة بالارقام العددية, مثل صفة لون العين (اسود , بني), الحالة الاجتماعية (غني , فقير) ...... الخ.

2. متغيرات كمية Quantitative Variables :

هي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بالارقام العددية مثل الطول, الوزن, السمك, التركيز, ... الخ. تقسم المتغيرات الكمية الى قسمين:

أ. متغيرات مستمرة Continuous Variables :

وهي المتغيرات التي تفترض وجود اي قيمة بين قيمتين او في مدى معين من القيم. مثال على ذلك ان عمر شخص (س) من الناس هو (50) سنة او (50.8) سنة أو (48.5) سنة , هذه القياسات او الارقام تعتمد على دقة التقدير لذلك تسمى متغيرات مستمرة.

ب. متغيرات منفصلة Discrete Variables :

وهي المتغيرات التي تفترض عدم وجود اي قيمة بين قيمتين , او لا يمكن قبول وجود قيمة معينة بين قيمتين. مثال على ذلك نفترض ان عدد اطفال عائلة معينة هو خمسة اطفال , ولكن لا يمكن ان نفترض ان عدد اطفال هذه العائلة هو 2.5 طفل او 4.5 طفل ..... وهكذا.

بصورة عامة يمكن القول ان القياسات او المتغيرات التي تستحصل بالقياس او التي يتم حسابها بالقياسات , هي متغيرات مستمرة لانها تقبل وجود الاعداد العشرية او الكسور مع الاعداد الصحيحة , بينما المتغيرات التي تستحصل بواسطة العدد او بمبدا العد فهي متغيرات منفصلة لانها لا تقبل الكسور العشرية.

الرموز الاحصائية Statistical Notation :

غالبا ما يتم استخدام رموز خاصة في العمليات الاحصائية المختلفة, وهي رموز عالمية تستخدم كما هي لعدم وجود مرادفات لها في اللغة العربية لحد الان, لغرض حل وانجاز العمليات الحسابية والمعادلات والقوانين الاحصائية . يرمز الى اي متغير بالرمز (y) ونعني بالمتغير (نموذج) , عندما يكون لدينا قيم عديدة للمتغير (y) في هذه الحالة نرمز لهذه المتغيرات العديدة بالرمز (yi) التي تمثل قيم او قراءات متعددة التي تستحصل من دراسة المتغير (Y) . مثال على ذلك لو كان لدينا اعمار خمسة طلاب وكما يلي:

yi = 20 , 18 , 22 , 23 , 15

هذا يعني ان :

20 = y1 = تمثل القيمة الاولى للمتغير الاول او النموذج الاول

18 = y2 = تمثل القيمة الثانية للمتغير الثاني او النموذج الثاني

22 = y3 = تمثل القيمة الثالثة للمتغير الثالث او النموذج الثالث

23 = y4 = تمثل القيمة الرابعة للمتغير الرابع او النموذج الرابع

15 = y5 = تمثل القيمة الخامسة للمتغير الخامس او النموذج الخامس

.... = yn = تمثل القيم التالية للمتغيرات المتبقية ان وجدت , الى ان نصل الى المتغير yn التي تطلق على القيمة الاخيرة المفترضة في المجموعة , وهي كما في المثال : yn=y5=15

يرمز الى مجموع عدد من القيم بالرمز ($\sum$) ويسمى سكما (Sigma) , ونعني به المجموع الاحصائي لعدد من القيم (Summation) .

يرمز الى مجموع عدد من القيم غير محددة لاي متغير بالرمز: ($\sum_{i=1}^{i=n} yi$) , وهذا يعني ان قيم النماذج تشتمل على النماذج من (1) وهو المتغير الاول الى قيمة النموذج الاخير وهو (n) . اي ان :

$$\sum_{i=1}^{n} yi = y1 + y2 + y3 + y4 + \cdots + yn$$

واحيانا يكتب للسهولة $\sum yi$ الذي يمثل مجموع قيم النماذج المطلوبة.

يرمز الى مجموع المجاميع الجزئية بالرمز $\sum_{i=3}^{5} yi$ وهذا يعني: ان هنالك جمع لجزء من الاعداد ضمن المجاميع الكلية وكما يلي:

$$\sum_{i=3}^{5} yi = y3 + y4 + y5$$

يرمز الى مجموع مربعات النماذج او المتغيرات بالرمز:

$$\sum_{i=1}^{n}(yi)^2$$

وهذا يعني ان :

$$\sum_{i=1}^{n}(yi)^2=(y1)^2+(y2)^2+(y3)^2+(y4)^2+\cdots+(yn)^2$$

يرمز الى مربع مجموع النماذج او المتغيرات بالرمز :-

$$\left(\sum_{i=1}^{n}yi\right)^2$$

وهذا يعني ان :-

$$\left(\sum_{i=1}^{n}yi\right)^2=(y1+y2+y3+\cdots+yn)^2$$

يرمز الى مجموع حاصل ضرب قيم متغيرين او نموذجين X و y بالرمز :

$$\sum_{i=1}^{n}Xi\ x\ yi$$

وهذا يعني ان :-

$$\sum_{i=1}^{n}Xi\ x\ yi\ =X1y1+X2y2+X3y3+\cdots+Xnyn$$

يرمز الى حاصل ضرب مجموعتين من النماذج $(\sum Xi)(\sum yi)$ وتساوي:-

$(\sum Xi)(\sum yi)= (X1+X2+X3+.....+Xn)(y1+y2+y3+.....+yn)$

اذا كانت *(C)* اي عدد ثابت فان:

$$\sum_{i=1}^{n}C=nc=C1+C2+C3+\cdots+Cn=nc$$

اذا كانت *(C)* اي عدد ثابت فان:

$$\sum_{i=1}^{n}Cyi=C\sum_{i=1}^{n}yi=Cy1+Cy2+Cy3+\cdots+Cyn=C(y1+y2+y3+\cdots+yn)$$
$$=C\sum y1$$

جمع قيم متغيرين او اكثر هو عبارة عن حاصل مجموع جمعهم: اي ان:-

$$\sum(Xi+yi)$$
$$=\sum Xi+\sum yi=(X1+y1)+(X2+y2)+\cdots+(Xn+yn)$$
$$=(X1+X2+X3+\cdots+Xn)+(y1+y2+y3+\cdots yn)$$
$$=\sum Xi+\sum yi$$

يرمز الى مجموع حاصل قسمة عددين يساوي :-

$$\sum\frac{Xi}{yi}=\ \frac{X1}{y1}+\ \frac{X2}{y2}+\cdots+\frac{Xn}{yn}=\frac{X1+X2+X3+\cdots+Xn}{y1+y2+y3+\cdots+yn}$$

<u>مسائل تطبيقية:</u>

1. اذا كان لدينا قيم كل من المتغيرات :

yi= 3 , 9 , 6 , 2 , Xi = 4 , 2 , 3 , 7

اوجد قيمة كل مما ياتي:

A. $\sum_{i=1}^{n} yi$ B. $\sum_{i=1}^{3} yi$ C. $\sum yi^2$

D. $[\sum yi]^2$ E. $\sum Xi\ yi$ F. $(\sum Xi)(\sum yi)$

<u>الحـــــل:-</u>

A. $\sum_{i=1}^{n} yi=y1+y2+y3+y4=3+9+6+2=20$

B. $\sum_{i=1}^{3} yi=y2+y3=9+6=15$

C. $\sum yi^2=\ y1^2+\ y2^2+\ y3^2+\ y4^2=(3)^2+(4)^2+(6)^2+(2)^2=130$

D. $[\sum yi]^2+(y1+y2+y3+y4)^2=(3+9+6+2)^2=(20)^2=\ 400$

E. $\sum Xi\ yi=X1y1+X2y2+X3y3+X4y4=4x3+2x9+3x6+7x2$ =62

F. $(\sum Xi)(\sum yi)=(X1+X2+X3+X4)(y1+y2+y3+y4)=16\ x\ 20=$
320

2. اذا كان لدينا قيم كل من النموذجين y , X وكما يلي:-

Xi = 2 , 6 , 3 , 1 , yi = 3 , 9 , 6 , 2

اوجـــــد قيمة كل مما يلي:-

A. $\sum(yi\ +Xi)^2$

B. $\sum(Xi-3)(yi-5)$

C. $\sum Xi\ yi^2$

D. $\sum(yi\ - 3)$

E. $\sum yi - 3$

F. $\sum \frac{Xi+2}{yi}$

G. $\sum \frac{(Xi+2)}{\sum yi}$

<u>الحـــــــــل:-</u>

A.

$$\sum (yi\ + Xi)^2 = (y1 - X1)^2 + (y2 - X2)^2 + (y3 - X3)^2 + (y4 - X4)^2$$
$$= (3-2)^2 + (9-6)^2 + (6-3)^2 + (2-1)^2 = 20$$

B.

$$\sum (Xi - 3)(yi - 5)$$
$$= (X1 - 3)(y1 - 5) + (X2 - 3)(y2 - 5) + (X3 - 3)(y3 - 5)$$
$$+ (X4 - 3)(y4 - 5)$$
$$= \ (2-3)(3-5)\ +\ (6-3)(9-5) + (3-3)(6-5)$$
$$+ (1-3)(2-5) = 20$$

C. $\sum(yi\ - 3) = \sum yi - \sum(3) = \sum yi - n(3) = \sum yi - (4)(3) = 20 - 12 = 8$

D. $\sum yi - 3 = 20 - 3 = 17$

E. $\sum \frac{Xi+2}{yi} = \frac{X1+2}{y1} + \frac{X2+2}{y2} + \frac{X3+2}{y3} + \frac{X4+2}{y4} = \frac{2+2}{3} + \frac{6+2}{9} + \frac{3+2}{6} + \frac{1+2}{2} = \frac{164}{36}$

F. $\sum \frac{(Xi+2)}{\sum yi} = \frac{\sum Xi+(n)(2)}{\sum yi} = \frac{12+8}{20} = 1$

الفصل الثاني

التوزيع التكراري للبيانات Frequency Distribution

عندما يتم جمع بيانات اولية (نماذج او متغيرات لدراسة ظاهرة معينة او وسط معين (Population) تعتبر بمثابة قراءات خام لا يمكن الاستفادة منها او فهمها وهي على هذه الصورة , لذلك لا بد من ترتيبها بصيغة وتصنيفها بصيغة معينة لغرض تسهيل مهمة فهمها ومعالجتها بصورة مبسطة. هذه المعالجات تكون اما ترتيبها عدديا بصورة تصاعدية او تنازلية او تقسيمها وتصنيفها الى فئات او مجاميع تسهل عملية توزيع هذه القراءات على هذه المجاميع وبالتالي تسهيل مهمة فهم طبيعة توزيع هذه القراءات او البيانات.

احد اهم المعالجات الاحصائية للنتائج هي تقسيم او توزيع القراءات الى فئات او مجموعات او اصناف (Classes) ثم يتم توزيع هذه القراءات على هذه الفئات او الاصناف. هذه العملية تسمى بالتوزيع التكراري للبيانات او العرض الجدولي للبيانات, والتي هي عبارة عن جداول تلخص فيها بيانات العينة الى فئات ولكل فئة تكرار تتحدد قيمته حسب البيانات. هناك نوعان من التوزيع التكراري للبيانات هما:-

أ. <u>التوزيع التكراري البسيط او الجدول البسيط:</u>

عبارة عن تكوين جدول يتكون من عمودين يتم توزيع البيانات او النتائج حسب صفة واحدة او متغير واحد على كل فئة . العمود الاول يتضمن تقسيمات الفئات او المجاميع او الاصناف, ام العمود الثاني يتضمن تقسيم التكرار او عدد البيانات (تكرار النتائج) او عدد القيم التي تقع ضمن كل فئة والتي تسمى بفئة التكرار Class (Frequency) , هذا التنظيم او الترتيب بهذه الصيغة يسمى التوزيع التكراري للبيانات (Frequency Distribution) او الجدول التكراري (Frequency Table) .

<u>مثال على ذلك:</u> اذا كان لدينا بيانات عن وزن (100) طالب من جامعة تكريت , وتم توزيع اوزان الطلبة الى فئات كما مدرجة في الجدول (2-1):

جدول (2-1) توزيع اوزان الطلبة

الفئات الوزن (كغم)	التكرار
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 - 74	8
المجموع الكلي	100

نلاحظ ان الفئة الاولى التي تتضمن مدى الوزن من (62 – 60) كغم تحتوي على خمسة طلاب فقط, والفئة الثانية التي تتضمن مدى الوزن من (65 – 63) كغم تحتوي على ثمانية عشر طالبا فقط, وهكذا يتم تقسيم بقية اعداد الطلبة على ضوء اوزانهم الى هذه الفئات لنحصل في النهاية على اعداد الطلبة موزعة حسب الوزن الى فئات او اصناف يتم من خلالها سهولة فهم ومعرفة طبيعة توزيع اوزان الطلبة وضمن اي فئة من الوزن تحتوي على اكبر عدد من الطلاب.

نتعرف على بعض المصطلحات التي يرد ذكرها ضمن تعريف الجدول التكراري ,هي:-

1. حجم العينة : هو عدد القراءات او البيانات التي تم قياسها او استحصالها من دراسة ظاهرة معينة, ويرمز لها بالرمز (N) في حالة البيانات غير المبوبة و $(\sum fi)$ في حالة البيانات المبوبة.
2. المدى Range : الفرق بين اعلى قيمة واقل قيمة بين البيانات او النتائج المستحصلة من دراسة ظاهرة معينة, ويرمز له بالرمز (R). ويحسب من العلاقة التالية: $\boldsymbol{R = max.value - min.value}$
3. الفئـــة Class:- هي عبارة عن مجاميع من القيم او اصناف من القيم يتم توزيع البيانات او القراءات على هذه المجاميع كل حسب قيمته . يتم تحديد عدد الفئات في اي جدول تكراري حسب العلاقة التالية:

$$\text{عدد الفئات} = \boldsymbol{1 + 3.3logN}$$

ملاحظة:(عادة يتم اختيار عدد الفئات لتكون ما بين (5) الى (10) فئات لسهولة بناء جدول تكراري وذلك اعتمادا على عدد القيم او البيانات المستخدمة, ومقدار الفروقات بين هذه القيم).

1. التكـرار:- يمثل عدد القيم او البيانات التي تقع ضمن كل فئة حسب القيم التي تقع بين الحد الاعلى والحد الادنى لكل فئة, ويرمز له بالرمز ($\mathbf{f_i}$).
2. طول الفئة (Class Interval) :- عادة يتم تحديد طول الفئة او مدى الفئة بقيمة عليا وقيمة دنيا , وبالعودة الى المثال السابق نجد ان الفئة الاولى تحدد بالقيمة 60 و 62 . هذه الفئة تشتمل على كافة القيم التي تقع بين هذين القيمتين. (يمكن ان نعرف طول الفئة بانه مقدار المدى بين حدي الفئة). يتم حساب قيمة طول الفئة من خلال عدة علاقات رياضية وكما يلي:-

A.

$$\text{طول الفئة} = \frac{\text{المدى (اعلى قيمة–اقل قيمة)}}{\text{عدد الفئات}}$$

B. طول الفئة = الحد الاعلى – الحد الادنى +1

C. طول الفئة = الحد الاعلى الحقيقي للفئة – الحد الادنى الحقيقي للفئة

D. طول الفئة هو الفرق بين الحدين الاعلى او الحدين الادنى لفئتين متتاليتين .

علما بان جميع اطوال الفئات يجب ان تكون متساوية.

3. حدود الفئة :- وهي القيم التي تقع في بداية ونهاية الفئة . كما في المثال السابق تكون القيم 62 و 60 تسمى حدود الفئة (Class Limit) , القيمة الدنيا التي هي 60 تسمى الحد الاسفل للفئة, والقيمة العليا التي هي 62 تسمى الحد الاعلى للفئة.

الفئات التي ليس لها حد اعلى او ليس لها حد ادنى تسمى الفئات المفتوحة , مثال على ذلك الفئات من عمر 40 سنة فما فوق او من عمر 20 سنة فما دون.

4. الحدود الحقيقية للفئات:- عندما يتم جمع بيانات او قيم تحتوي على كسور عشرية او ارقام عددية بين الارقام الرئيسية (بيانات مستمرة) , مثلا جمع نتائج وزن عدد من الطلاب تحتوي على كسور, بهذه الحالة فان الفئة (62 – 60) تحتوي او تضم الاعداد من 59.5 الى 62.5 من حدود الكسر العشري الادنى الى حدود الكسر العشري الاعلى , هذه الحدود هي التي تسمى الحدود الحقيقية للفئات. لكل فئة يوجد حد حقيقي اعلى وحد حقيقي ادنى . يتم اللجوء الى تحديد الحدود الحقيقية للفئات لغرض تجاوز التداخل او الغموض في القراءات اثناء توزيعها على الفئات خاصة في حالة وجود كسور عشرية. كما في المثال السابق فان الحد الاعلى الحقيقي للفئة الاولى يساوي 62.5 اما الحد الادنى الحقيقي لها يساوي 59.5.

كيفية استخراج الحد الحقيقي الاعلى للفئة والد الادنى للفئة :-

1. حاصل جمع الحد الاعلى للفئة الاولى مع الحد الادنى للفئة التالية لها او التي ياتي بعدها مقسوماً على (2), مثلا كما في المثال السابق , ان الحد الاعلى للفئة الاولى هو (62) والحد الادنى للفئة الثانية هو (63) فيكون الحد الحقيقي الاعلى للفئة الاولى هو $(62.5 = \frac{62+63}{2})$.

2. الطريقة الثانية في الحساب هي :

الحد الادنى الحقيقي لاي فئة = مركز تلك الفئة – ½ (طول الفئة)

الحد الاعلى الحقيقي لاي فئة = مركز تلك الفئة + ½ (طول الفئة)

3. الحد الادنى الحقيقي لاي فئة = (الحد الادنى لتلك الفئة – 0.5), والحد الاعلى الحقيقي لاي فئة = (الحد الاعلى لتلك الفئة + 0.5).

واحيانا تكتب الفئات بالاعتماد على الحدود الحقيقية العليا والدنيا وكالآتي:

التكرار	الفئات

59.5 – 62.5	5
62.5 – 65.5	18
65.5 – 68.5	42
68.5 – 71.5	27
71.5 – 74.5	8

من فوائد اختيار الحدود الحقيقية للفئات هي تجاوز الغموض الذي يحصل عند توزيع البيانات على الفئات وعدم معرفة هل ان بعض القراءات تعود الى الفئة الاقلى مثلا, او الفئة الثانية. عند اختيار الحدود الحقيقية للفئات يتم تاكيد توزيع البيانات الحاوية على كسور الى الفئات بصورة صحيحة واكيدة.

1. مركز الفئة أو متوسط الفئة (Class Midpoint) :

تعرف مركز الفئة بانها تلك القيمة التي تقع في منتصف الفئة أو قيمة متوسط الفئة. وتحسب قيمتها كما يلي:-

$$\text{مركز الفئة} = \frac{\text{الحد الاعلى للفئة} + \text{الحد الادنى للفئة}}{2}$$

مركز الفئة من المثال السابق = $\frac{60+62}{2} = 61$

من فوائد حساب مركز الفئة هي امكانية اختزال كافة القيم التي تقع ضمن كل فئة الى قيمة واحدة فقط, اي انه في حالة وجود مزيد من القيم ضمن الفئة الواحدة فيمكن اعتبارها جميعا تساوي قيمة واحدة فقط قيمتها هي مركز الفئـــــة.

كيفية تكوين او بناء جدول تكراري

1. تحديد اعلى قيمة واقل قيمة في البيانات او القراءات المستحصلة من دراسة ظاهرة معينة.
2. ايجاد المدى (Range) والذي يمثل الفرق بين اعلى قيمة واقل قيمة ضمن البيانات المستحصلة.
3. تقسيم المدى الى عدد مناسب من (اطوال الفئات) متساوية الاطوال تضم كافة قيم البيانات. ويمكن استخدام العلاقة السابقة:

عدد الفئات $= 1 + 3.3logN$ (اي ان عدد او حجم البيانات ومقدار المدى هو الذي يحدد عدد الفئات) .
علما بان عدد الفئات المتعارف عليه في المعالجات الاحصائية تقع بين 5 و 10 فئات .

4. حساب طول الفئة الذي يساوي :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

5. يتم توزيع البيانات او القراءات على هذه الفئات لمعرفة عدد القيم او القراءات التي تقع ضمن كل فئة. ولتسهيل مهمة حساب عدد القيم التي تقع ضمن كل فئة يتم ذلك بتسجيل القيم بشكل اشارات او علامات ثم يتم جمعها وتحويلها الى رقم محدد, ويتم تنظيم هذه العملية في بعض الاحيان ضمن جدول ابتدائي يسمى (جدول العلامات او جدول تفريغي) يتم فيه تسجيل البيانات التي تقع ضمن كل فئة على شكل علامات كما في الجدول (2−2):

جدول (2−2) جدول تفريغي وجدول تكراري

جدول العلامات او الجدول التفريغي ⟵ الجدول التكراري

الفئات الوزن (كغم)	العلامات	التكرار
60 – 62	~~////~~	5
63 – 65	~~////~~ ~~////~~ ~~////~~ ///	18
66 – 68	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ //	42
69 – 71	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ //	27
72 - 74	~~////~~ ///	8
المجموع الكلي	100	100

الفئات الوزن (كغم)	التكرار
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 - 74	8
المجموع الكلي	100

6. تنظم هذه العملية في جدول متكون من حقلين , حقل الى الفئات والحقل الاخر الى التكرار الذي يمثل العدد التكراري للقيم الواقعة ضمن كل فئة.
7. يجب ان نبدأ بكتابة الحد الادنى للفئة الاولى بقيمة اقل من اصغر قيمة موجودة ضمن البيانات او القراءات, وتنتهي الفئات بالحد الاعلى للفئة الاخيرة بقيمة اكبر بقليل من اعلى قيمة موجودة ضمن البيانات او القراءات الاحصائية.

ب. الجدول المركب أو التوزيع التكراري المركب:

وهو الجدول الذي توزع فيه البيانات او النتائج حسب صفتين أو ظاهرتين في نفس الوقت. مثلا تكون الصفوف الافقية تمثل احدى الصفات والصفوف العمودية تمثل الصفة الثانية , اما المربعات او المساحات التي تمثل تقاطع هذين الفئتين فإنها تحتوي على عدد التكرارات او الصفات المشتركة .جدول (2−3) يمثل جدول تكراري مركب لتوزيع اوزان (100) طالب حسب الطول:

جدول (2−3) جدول تكراري مركب

الوزن / الطول	51 – 60	61 – 70	71 – 80	المجموع
140 – 121	20	6	4	30
160 – 141	2	40	10	52
180 – 161	2	6	10	18
المجموع	24	52	24	100

<u>التوزيع التكراري التجمعي (التصاعدي أو التنازلي) (Cumulative Frequency Distribution):</u>

يعرف التوزيع التكراري التجمعي بانه حاصل جمع كافة التكرارات للبيانات ابتداءً من تكرار الفئة الاولى صعودا لغاية تكرار الفئة الاخيرة أو بالعكس ابتداءً من تكرار الفئة الاخيرة نزولاً الى تكرار الفئة الاولى.

كما في الجدول (2–4):

جدول (2–4) جدول تكراري تجمعي

الفئات	التكرار	التكرار التجمعي الصاعد	التكرار التجمعي النازل
60 – 62	5	5	100
63 – 65	18	23	95
66 – 68	42	65	77
69 – 71	27	92	35
72 - 74	8	100	8
المجموع الكلي	100		

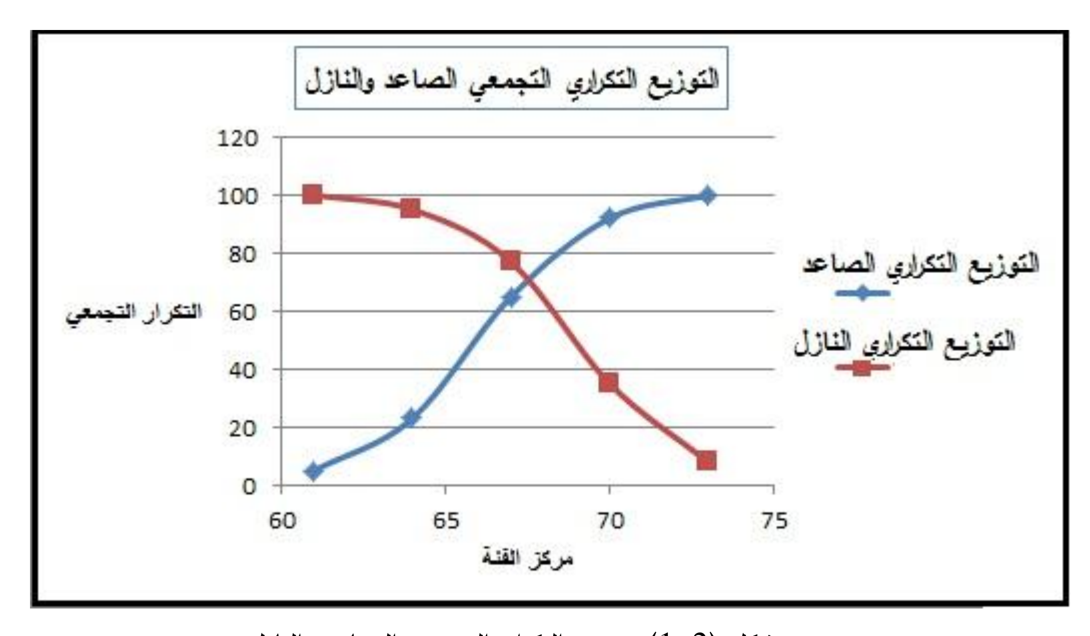
الجدول الذي يحتوي على مثل هذه التكرارت المتجمعة يسمى التوزيع التكراري المتجمع, اذا كان تصاعديا يسمى التكرار التجمعي التصاعدي واذا كان تنازليا يسمى التكرار التجمعي التنازلي, او باختصار يسمى التوزيع التكراري التجمعي. احيانا يتم تنظيم جدول تكراري تجمعي باستخدام الحدود الحقيقية العليا للفئات كما في الجدول (2–5):

جدول (2-5) جدول الحدود الحقيقية للفئات

الحدود الحقيقية العليا الفئات	التكرار	التكرار التجمعي الصاعد	التكرار التجمعي النازل
Less than 62.5	5	5	100
Less than 65.5	18	23	95
Less than 68.5	42	65	77
Less than 71.5	27	92	35
Less than 74.5	8	100	8
المجموع الكلي	100		

<u>التمثيل البياني الى قيم التكرار التجمعي:</u>

يمكن عمل او رسم تمثيل بياني الى قيم التكرار التجمعي الصاعد والتكرار التجمعي النازل والذي يكون على شكل منحني يسمى (Ogive) او على شكل S – Shape كما في الشكل (2-1):



شكل (2-1) منحني التكرار التجمعي الصاعد والنازل

التوزيع التكراري النسبي Relative frequency Distribution :

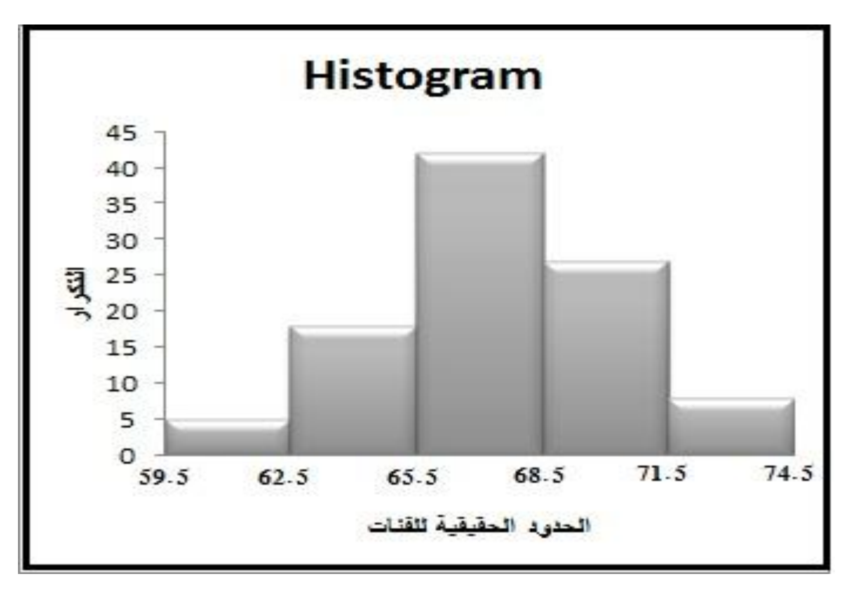
يعرف التوزيع التكراري النسبي لاي فئة بانه التكرار لتلك الفئة مقسوماً على التكرار الكلي لمجموع الفئات, وعادة يكتب بصيغة نسبة مئوية (%) , حيث يتم ضرب حاصل القسمة في (100). مثال على ذلك اذا كان التكرار النسبي لاحد الفئات هو 42 ومجموع التكرارات لجميع الفئات يساوي (100) فان التكرار النسبي لتلك الفئة يساوي: $\frac{42}{100}\ x\ 100 = \ 42\%$ وهكذا يحسب لبقية الفئات.

التوزيع التكراري التجمعي النسبي Relative Cumulative Frequency Distribution :

وهو يمثل التكرار التجمعي مقسوما على مجموع التكرارات للفئات كافة. مثال على ذلك اذا كان التكرار التجمعي لاحدى الفئات يساوي 35 وان مجموع التكرارات للفئات كافة يساوي 100 فان التكرار التجمعي النسبي لتلك الفئة يساوي: $\frac{35}{100}\ x\ 100 = \ 35\%$.

المدرج التكراري Histogram :

إن وسائل واساليب التمثيل البياني والصوري لعرض النتائج الاحصائية كثيرة ومتنوعة, وهي تمثل طريقة سهلة وفعالة تساعد على سرعة فهم واستيعاب البيانات وبناء استنتاجات واستخراج معلومات منها. أهم هذه الوسائل ومن اكثرها شيوعا هو المدرج التكراري Histogram , وهو عبارة عن مستطيلات راسية متجاورة تمتد قواعدها على المحور الافقي او المحور السيني. الذي يمثل طول الفئة او الحدود الحقيقية للفئات , اما طول او ارتفاع هذه المستطيلات فانها تمثل تكرار الفئات التي ترسم على المحور العمودي او المحور الصادي. كما في الشكل (2-2):



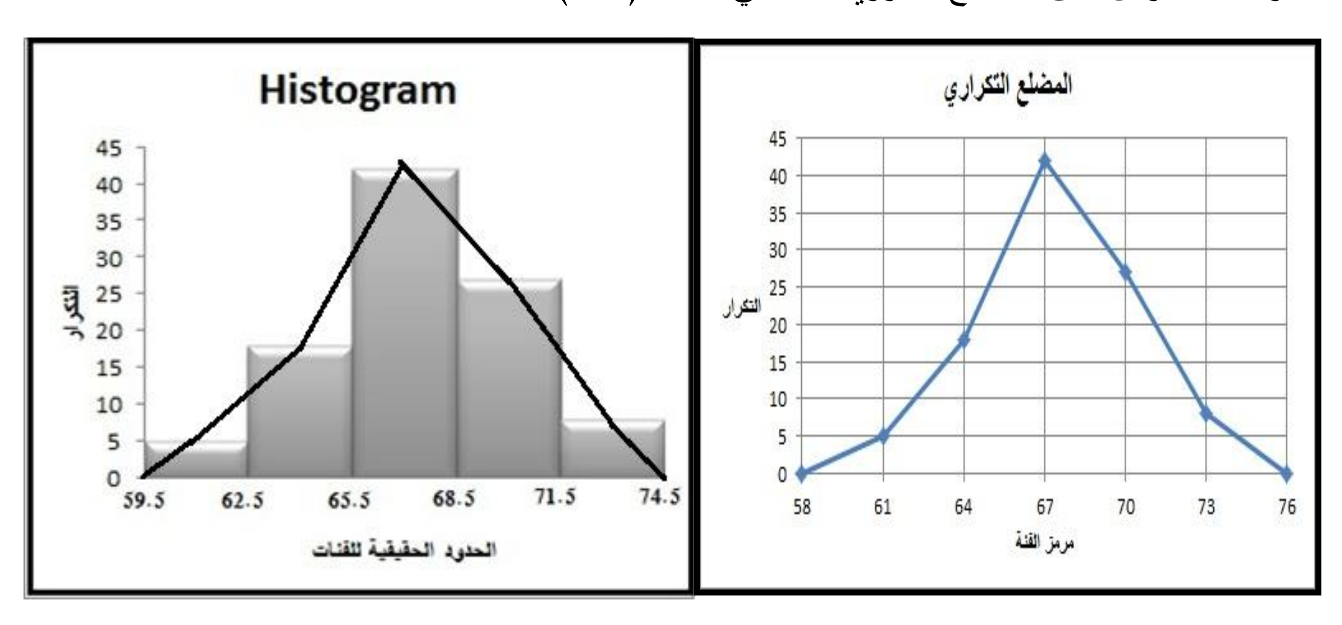
شكل (2-2) المدرج التكراري

كيفية بناء المدرج التكراري:

1. يرسم المحور الافقي الذي يمثل الحدود الحقيقية للفئات والمحور العمودي الذي يمثل التكرار (f_i) وفق مقياس رسم مناسب .
2. تؤشر الحدود الحقيقية للفئات او مراكز الفئات على المحور الافقي , ويفضل ان تترك مسافة مناسبة من نقطة الصفر لغاية بداية الحد الحقيقي الادنى للفئة الاولى لضمان عدم حصول تداخل بين المستطيل الاول والمحور العمودي.
3. يتم تاشير تكرار الفئات على المحور الصادي او المحور العمودي حيث تقسم الى مسافات متساوية.
4. يرسم ضمن كل فئة مستطيل راسي تمثل قاعدته طول الفئة وارتفاعه يمثل تكرار الفئة.

المضلع التكراري:

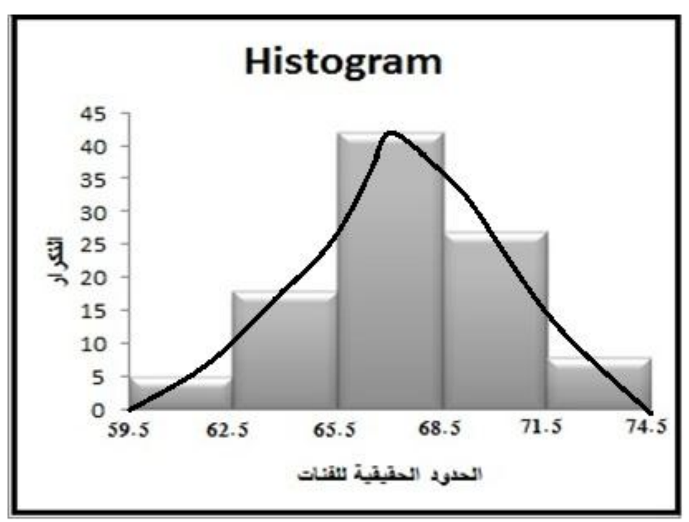
عند ربط النقاط التي تمثل منتصف الفئة التي تقع على راس المستطيل ينتج لنا منحني يسمى المضلع التكراري . يمكن تعريف المضلع التكراري بانه عبارة عن مستقيمات متصلة تصل بين مراكز الفئات وقيمة التكرار للبيانات المستحصلة من الجدول التكراري, ويفضل ان تبدأ من القيمة صفر على المحور السيني وتنتهي بالقيمة صفر كذلك لغرض غلق المضلع التكراري, كما في الشكل (2-3):

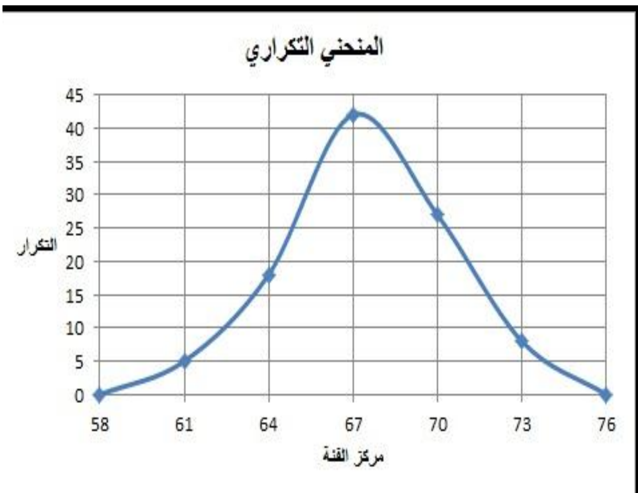


شكل (2-3) المضلع التكراري

المنحني التكراري:

هو عبارة عن خط منحني مهذب يربط بين نقاط مراكز الفئات والتكرار للبيانات في الجدول التكراري. يرسم بنفس طريقة المضلع التكراري ماعدا ان الخط يكون املس ومهذب ولا توجد فيه تكسرات, كما في الشكل (2−4):





شكل (2−4) المنحني التكراري

ويمكن ان يرسم المضلع التكراري والمنحني التكراري مباشرة على المحور الافقي الذي يمثل مراكز الفئات والمحور العمودي الذي يمثل التكرار دون الاعتماد على المدرج التكراري , حيث يتم تاشير النقاط الواصلة بين مركز الفئة والتكرار.

مسائل تطبيقية عن التوزيع التكراري للبيانات

مسألة (1):

البيانات المدرجة في الجدول (2−5) , تمثل قيم التحاليل الكيميائية الى (SiO_2) , (AL_2O_3) , (Fe_2O_3) في احد الابار اللبابية ,لغرض التحري عن ترسبات البوكسايت. أرسم واحسب:

أ. الجدول التكراري لتلك التحاليل الثلاثة لمعرفة طبيعة توزيع هذه المركبات.

ب. المضلع التكرار, ثم المنحني التكراري.

ت. التوزيع التكراري النسبي , ثم التوزيع التجمعي النسبي الصاعد والنازل.

جدول (2−5) بيانات التحاليل الكيميائية

العمق(m)	Fe2O3 %	Al2O3 %	SiO2 %	العمق(m)	Fe2O3 %	Al2O3 %	SiO2 %	العمق(m)	Fe2O3 %	Al2O3 %	SiO2 %
0.00	24.80	45.01	3.50	10.00	34.40	39.02	1.87	20.00	17.80	49.66	2.72
0.50	22.40	48.69	0.69	10.50	22.80	43.45	2.29	20.50	24.60	45.49	1.86
1.00	19.80	49.93	1.28	11.00	26.60	42.58	3.91	21.00	21.40	48.03	2.32
1.50	27.80	45.17	0.89	11.50	32.40	38.72	3.34	21.50	24.00	45.49	2.30
2.00	25.80	43.77	0.49	12.00	27.40	42.92	4.34	22.00	26.00	44.05	2.34
2.50	23.00	48.69	0.52	12.50	25.80	43.89	3.20	22.50	19.00	49.02	2.95
3.00	34.40	39.38	0.80	13.00	31.40	38.13	3.74	23.00	24.00	32.34	2.75
3.50	29.60	42.26	1.33	13.50	26.80	42.92	3.04	23.50	15.60	41.90	2.04
4.00	30.80	32.40	1.19	14.00	24.40	46.13	1.68	24.00	19.20	47.11	4.63
4.50	35.80	33.53	1.18	14.50	25.60	44.85	2.42	24.50	17.80	48.70	3.07
5.00	40.40	31.74	1.77	15.00	23.80	44.85	3.33	25.00	21.20	46.77	3.07
5.50	40.00	32.68	1.32	15.50	21.60	47.41	2.17	25.50	30.00	49.39	4.10
6.00	15.20	32.56	1.69	16.00	27.60	44.53	3.22	26.00	23.60	45.01	4.17
6.50	17.60	49.98	2.18	16.50	19.80	48.05	2.70	26.50	25.20	42.62	5.21
7.00	24.20	45.83	2.90	17.00	27.80	44.21	2.76	27.00	17.20	50.20	3.25
7.50	19.40	4.69	3.30	17.50	21.20	48.70	2.11	27.50	20.60	48.13	4.73
8.00	35.30	33.95	2.11	18.00	19.00	49.34	2.34	28.00	25.00	43.23	5.34
8.50	31.40	39.02	2.00	18.50	28.60	44.21	1.23	28.50	18.00	56.88	4.55
9.00	28.21	32.52	1.99	19.00	24.80	45.49	2.14	29.00	25.00	44.20	3.54
9.50	34.00	39.02	2.11	19.50	23.40	46.99	1.59	29.50	21.40	46.77	3.99

الحـــــل:

لغرض تكوين جدول تكراري , نهمل العمق لعدم اهميته في الجدول لانه لا يمثل احد المتغيرات المطلوبة, نهتم فقط بالمتغيرات في التحاليل الكيميائية التي هي (SiO_2) , (AL_2O_3) , (Fe_2O_3) .

نبدأ باول متغير الذي هو الهيماتايت (Fe_2O_3).

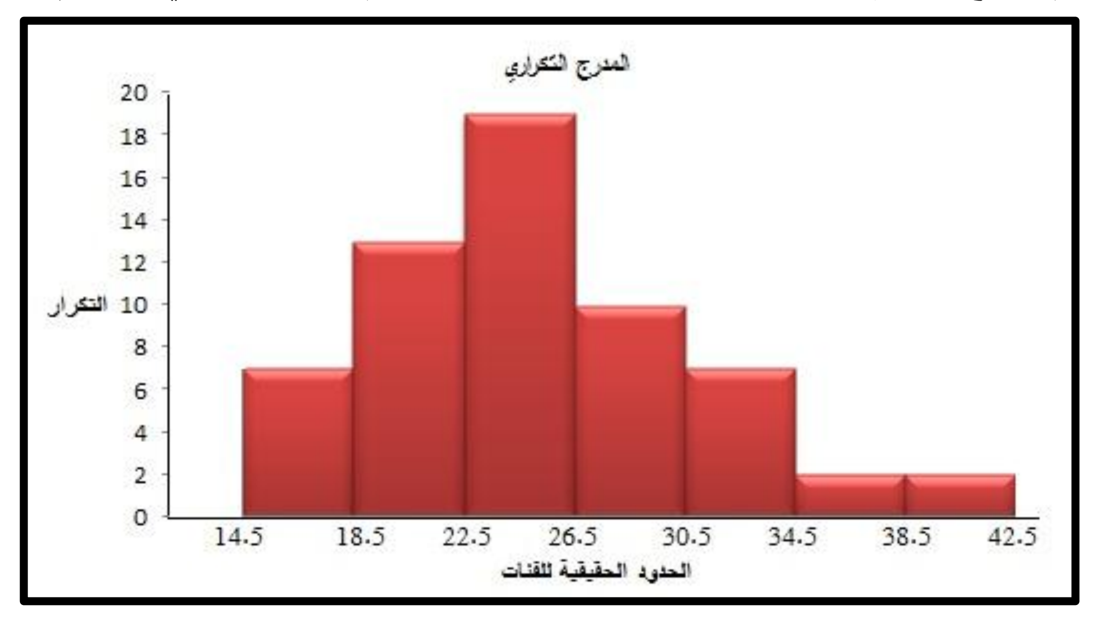
أ. نحسب ونرسم الجدول التكراري

1. نستخرج المدى الذي يمثل الفرق بين اعلى واقل قيمة في التحاليل الكيميائية= **40.4−15.2=25.2**
2. نختار عدد الفئات من العلاقة : عدد الفئات = $1 + 3.3logN = 1 + 3.3log60 = 6.8 \cong 7$
3. نستخرج طول الفئة والتي تساوي $\frac{\text{المدى}}{\text{عدد الفئات}} = \frac{25.2}{7} = 3.6 \cong 4$
4. نكون الجدول التكراري كما في الجدول (2−6), ونكتب عدد الفئات ابتداءً من حدود الفئات الدنيا ولغاية الحد الاعلى للفئة الاخيرة, ونظراً لوجود كسور عشرية في البيانات المعطاة, فعليه يجب استخدام الحدود الحقيقية للفئات واعتمادها في توزيع البيانات على الفئات.
5. نبدأ بتوزيع البيانات على الفئات باستخدام مبدأ العد او التأشير (Tally Marks) كما في لعبة كرة السلة.
6. بعد الانتهاء من التوزيع, نعد عدد التاشيرات ونكتب المجموع مقابل كل فئة الذي يمثل التكرار.

جدول (2−6) بناء جدول تكراري

الفئات	الحدود الحقيقية للفئات	التاشير	التكرار
15 − 18	14.5 − 18.5	~~1111~~ 11	7
19 − 22	18.5 − 22.5	~~1111~~ ~~1111~~ 111	13
23 − 26	22.5 − 26.5	~~1111~~ ~~1111~~ ~~1111~~ 1111	19
27 − 30	26.5 − 30.5	~~1111~~ ~~1111~~	10
31 − 34	30.5 − 34.5	~~1111~~ 11	7
35 − 38	34.5 − 38.5	11	2
39 − 42	38.5 − 42.5	11	2

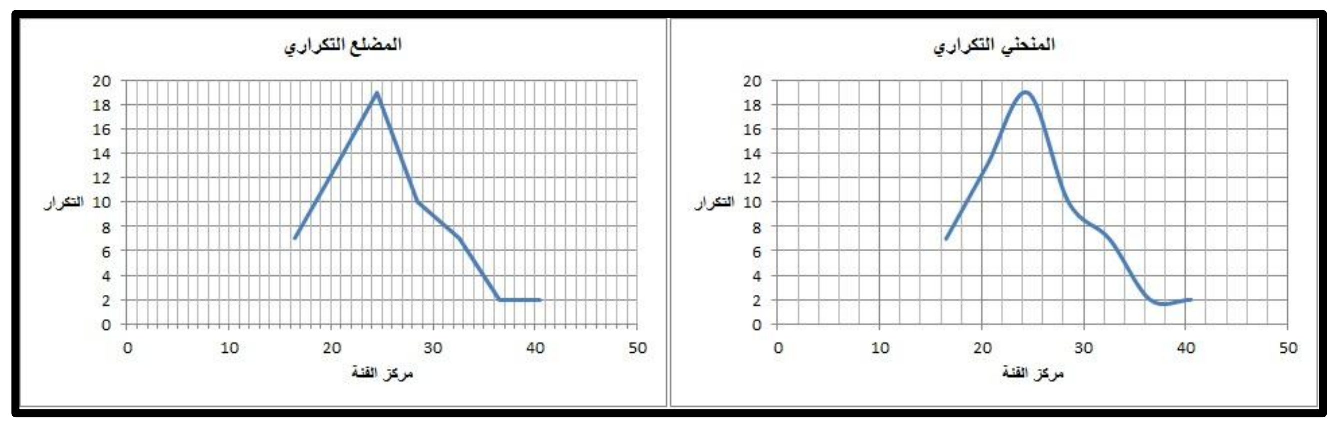
7. نرسم المدرج التكراري على شكل اعمدة متصلة يمثل ارتفاعها قيم التكرار , كما في الشكل (2−5).



شكل(2−5) رسم المدرج التكراري

ب. نرسم المضلع التكراري والمنحني التكراري:

يتم ايصال نقاط مراكز الفئات بواسطة خطوط مستقيمة لينتج منها مضلع تكراري, كما في الشكل (2−8), وعندما يتم عمل تهذيب لهذه الخطوط المستقيمة نحصل على المنحني التكراري, كما في الشكل (2−9).



شكل (2−9) المنحني التكراري

شكل (2−8) المضلع التكراري

ت. نحسب التوزيع النسبي:

نحسب التوزيع التكراري النسبي للفئات بواسطة قسمة تكرار كل فئة على مجموع التكرارات مضروباً في مئة, لنحصل على النسبة المئوية للتوزيع التكراري لكل فئة , كما في الجدول (2−8).

جدول (2–8) التوزيع التكراري النسبي

الفئات	مركز الفئة	التكرار	التكرار النسبي	التكرار التجمعي الصاعد	التكرار التجمعي النسبي
15 – 18	16.5	7	11.666	7	11.666
19 – 22	20.5	13	21.666	20	33.332
23 – 26	24.5	19	31.666	39	64.998
27 – 30	28.5	10	16.666	49	81.664
31 – 34	32.5	7	11.666	56	93.333
35 – 38	36.5	2	3.333	58	96.6633
42 - 39	40.5	2	3.333	60	100.00

وهكذا وبنفس الطريقة يتم استخراج ورسم بقية المتغيرات .

مسألة (2):

تم تسجيل الدرجات النهائية الى (70) طالب من طلاب المرحلة الثانية في مادة الاحصاء للعام الدراسي الحالي , كما مدونة في الجدول (2–9).

جدول (2–9) درجات الاحصاء لطلبة المرحلة الثانية

68	84	75	82	68	90	62	88	76	93
73	79	88	73	90	93	71	59	58	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

المطلـــــوب:

1. تنظيم جدول تكراري للبيانات.
2. رسم المدرج التكراري والمنحني التكراري.
3. رسم المنحني التكراري الصاعد والنازل.
4. من المنحني التكراري التجمعي الصاعد اوجد عدد الطلاب الذين تقل درجاتهم عن 80% والحد الاعلى للدرجة التي حصل عليها 50 طالب.

الحـــــــــل:

1. يتم تنظيم الجدول التكراري كما في الجدول (1−10) .

المدى = اعلى قيمة − اقل قيمة = 97 − 53 = 44

عدد الفئات = 1+3.33logN = 1+3.33log70 = 1+3.33 x 1.84 = 7.127= 8

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{\mathbf{44}}{\mathbf{8}} = \mathbf{5.5} \cong \mathbf{6}$$

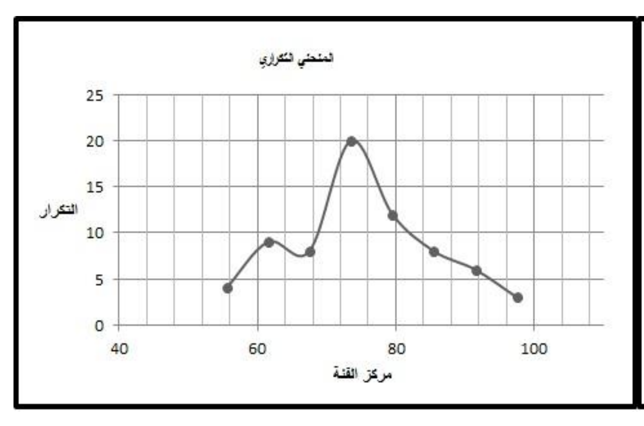
جدول (2−10) تنظيم الجدول التكراري

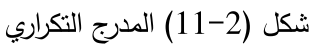
الفئات	التكرار	الحدود الحقيقية للفئات	مركز الفئة	التكرار التجمعي الصاعد	التكرار التجمعي النازل
53 – 58	4	52.5 – 58.5	55.5	4	70
59 – 64	9	58.5 – 64.5	61.5	13	66
65 –70	8	64.5 – 70.5	67.5	21	57
71 – 76	20	70.5 – 76.5	73.5	41	49
77 – 82	12	76.5 –82.5	79.5	53	29
83 – 88	8	82.5 – 88.5	85.5	61	17
89 – 94	6	88.5 – 94.5	91.5	67	9
95 – 100	3	94.5 – 100.5	97.5	70	3

2. نرسم المدرج التكراري مع المنحني التكراري اعتمادا على البيانات في الجدول التكراري (2−10) كما في الشكلين (2−10) و (2−11) على التوالي.

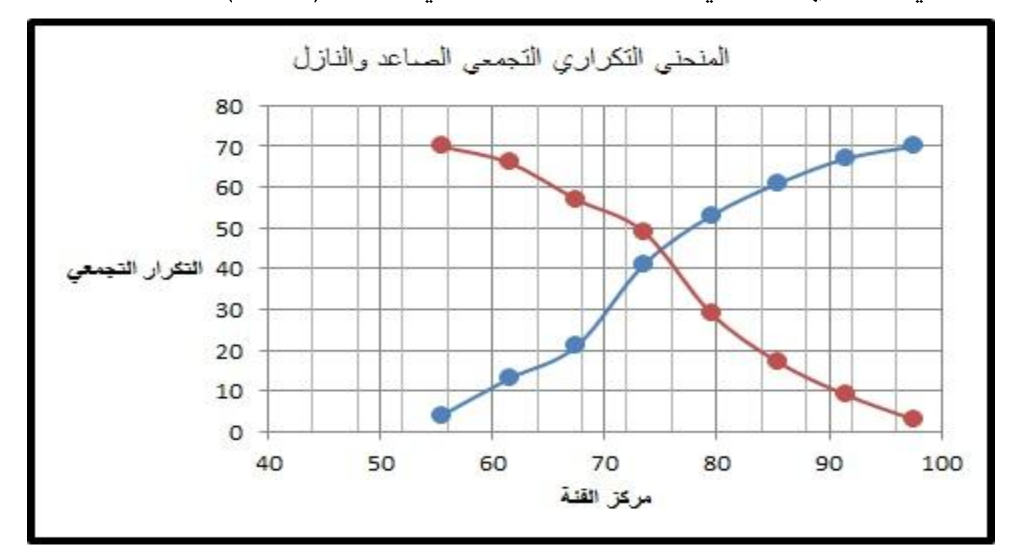


شكل (2−10) المنحني التكرري



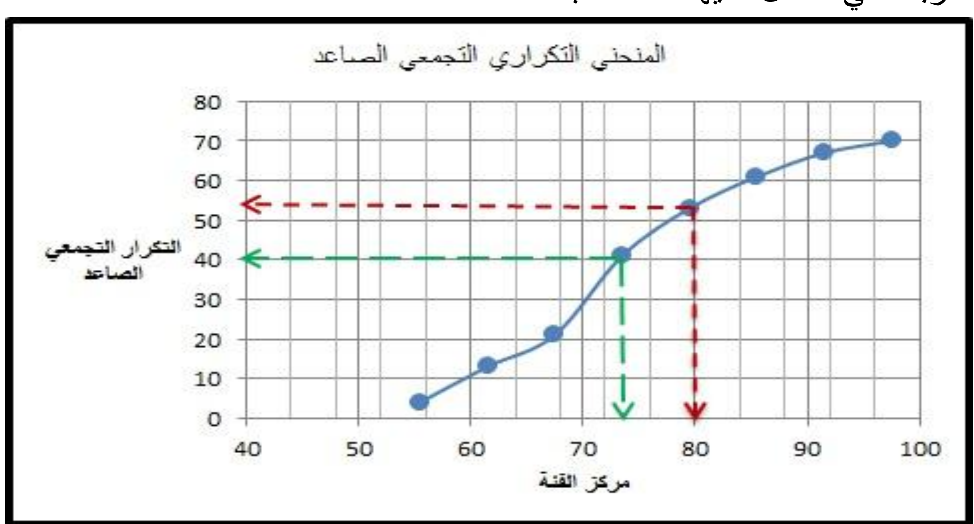

شكل (2−11) المدرج التكراري

3. نرسم المنحني التكراري التجمعي الصاعد والنازل كما في الشكل (2−12)



شكل (2−12) شكل المنحني التكراري الصاعد والنازل

4. نرسم المنحني التكراري التجمعي الصاعد ثم نوجد عدد الطلاب الذين تقل درجاتهم عن 80% والحد الاعلى للدرجة التي حصل عليها 40 طالب.



من المنحني نستخرج بواسطة رسم السهم الاحمر نلاحظ ان عدد الطلاب الذين حصلوا على 80% من الدرجة يبلغ عددهم (54) طالب , وكذلك هناك (40) طالب حسب ما مؤشر بالسهم الاخضر هم الذين حصلوا على 74% من الدرجة في مادة الاحصاء.

مسألة (3):

تم اجراء تحليل كيميائي الى (54) نموذج من ترسبات تكوين رملي في منطقة الصحراء الغربية لمعرفة تراكيز السيليكا (SiO_2) وكانت النتائج كما مدرجة في الجدول (2−11):

جدول (2−11) بيانات التحاليل الكيميائية

رقم النموذج	(SiO2) %	رقم النموذج	(SiO2) %	رقم النموذج	(SiO2) %	رقم النموذج	(SiO2) %	رقم النموذج	(SiO2) %	رقم النموذج	(SiO2) %
S1	40	S10	54	S19	58	S28	57	S37	52	S46	45
S2	41	S11	35	S20	56	S29	48	S38	60	S47	41
S3	44	S12	48	S21	52	S30	54	S39	39	S48	47
S4	51	S13	55	S22	46	S31	36	S40	51	S49	50
S5	42	S14	43	S23	48	S32	37	S41	38	S50	42
S6	48	S15	55	S24	31	S33	53	S42	41	S51	47
S7	50	S16	45	S25	47	S34	42	S43	47	S52	50
S8	46	S17	48	S26	48	S35	43	S44	48	S53	60
S9	47	S18	47	S27	54	S36	46	S45	49	S54	44

المطلوب:−

A. انشاء جدول تكراري باستعمال ستة فئات.

B. ارسم المدرج التكراري.

C. ارسم المنحني التكراري.

D. ارسم المنحني التكراري الصاعد والنازل.

الحـــــــــــل:-

A. يتم تنظيم الجدول التكراري باستخدام ستة فئات كما في الجدول (2−12).

المدى = اعلى قيمة – اقل قيمة = 60 − 31 = 29

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{29}{6} = 4.5 \cong 5$$

جدول (2−12)

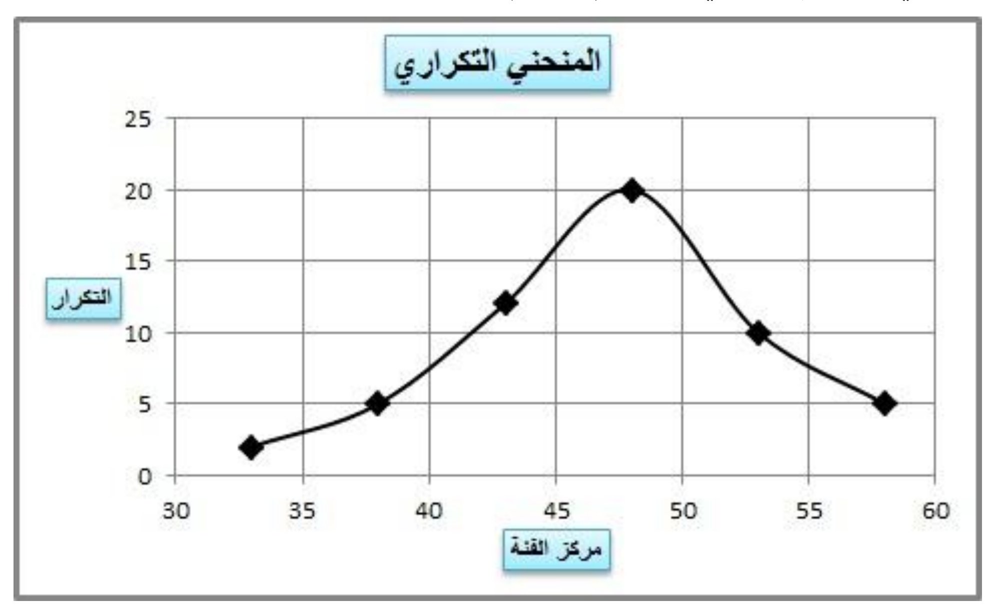
الفئات	التكرار	مركز الفئة	التكرار التجمعي الصاعد	التكرار التجمعي النازل
31 – 35	2	33	2	54
36 – 40	5	38	7	52
41 – 45	12	43	19	47
46 – 50	20	48	39	35
51 – 55	10	53	49	15
56 - 60	5	58	54	5

B. نرسم المدرج التكراري وكما في الشكل (2−13):



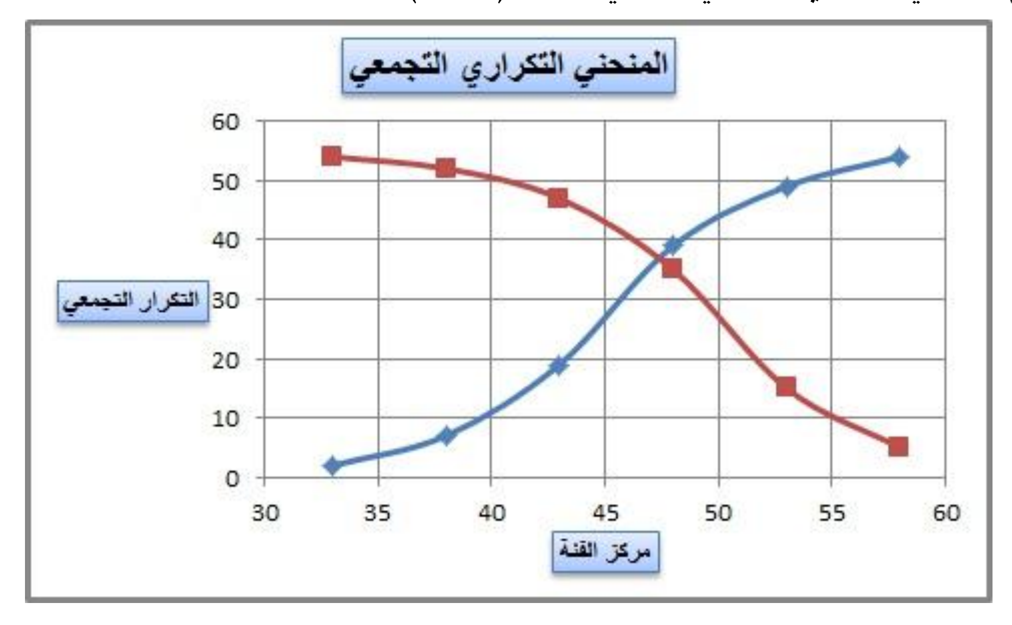
شكل (2−13) المدرج التكراري

C. نرسم المنحني التكراري كما في الشكل (2−14):



شكل (2−14) المنحني التكراري

D. نرسم المنحني التكراري التجمعي كما في الشكل (2−15):

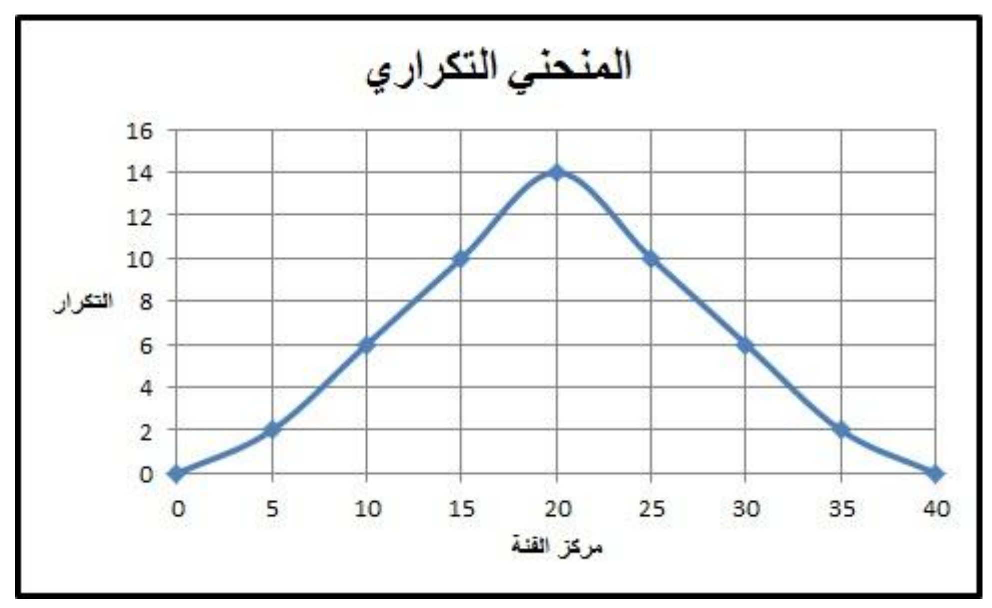


شكل (2−15) المنحني التكراري التجمعي

الفصل الثالث

المنحني التكراري Frequency Curve :

هو عبارة عن منحني يمر بكافة النقاط الواقعة في مراكز الفئات والتي يمثل ارتفاعها تكرارات تلك الفئات. يمكن الحصول على منحني تكراري من المضلع التكراري (Polygon or the Ogive) بعد اجراء عملية تهذيب (Smoothing) الى الخطوط المتكسرة, كما في الشكل (3−1):



شكل (3−1) شكل المنحني التكراري

عادة يتم اقفال المنحني التكراري في البداية والنهاية وذلك بان توصل بدايته بالحد الادنى للفئة الاولى ونهايته بالحد الاعلى للفئة الاخيرة. مساحة المنطقة الواقعة تحت المنحني مكافئة (وليست مساوية) لمساحة المدرج التكراري.

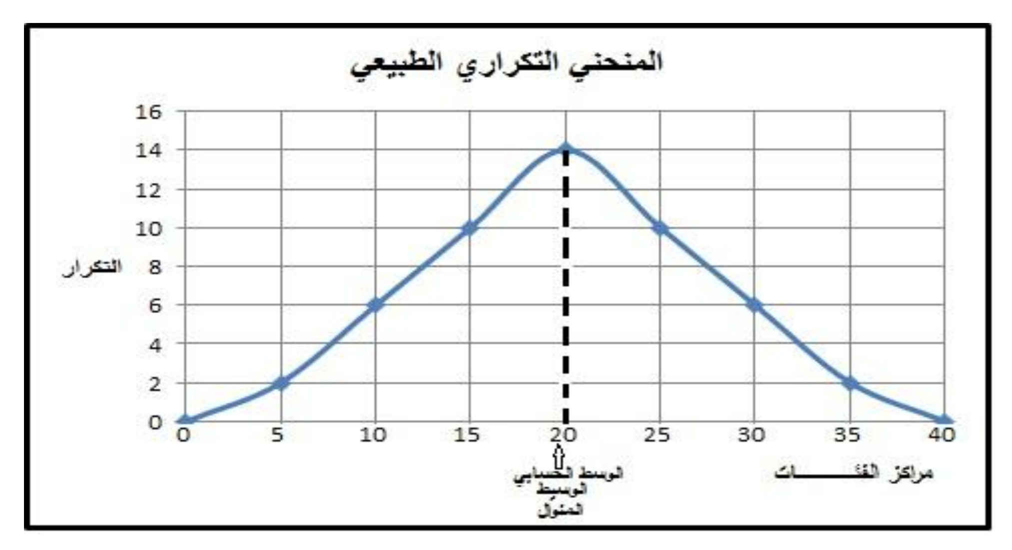
انواع المنحنيات التكرارية Types of Frequency Curves :

اهم الاشكال للمنحنيات التكرارية التي من الممكن ان نحصل عليها اثناء عمليات المعالجة الاحصائية للنتائج من عملية رسم المنحني التكراري هي:−

a) المنحني الطبيعي Symmetrical Shape :

يسمى المنحني الطبيعي كذلك بمنحني كاوس (Gaussian Curve) نسبة الى العالم الذي اشتق معادلة المنحني عام (1777−1865) , واحيانا يسمى منحني كاوس − لابلاس.

وهو المنحني الذي يتصف بان القيم او البيانات تتوزع بشكل متماثل على جانبي خط المنتصف, أو على يمين ويسار الوسط الحسابي ($\bar{X}$) Mean للقراءات او البيانات . إن نهايات المنحني من جهة اليمين واليسار تمثل التكرارات القليلة للقيم في الفئات العليا والدنيا, كما في الشكل (3−2):



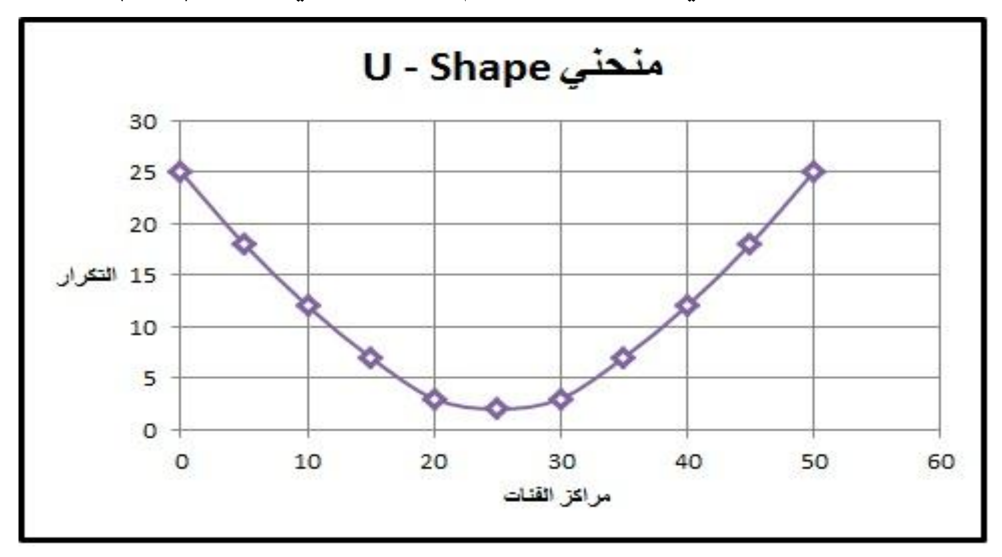
شكل (3−2) المنحني التكراري الطبيعي

خواص المنحني الطبيعي:

1. شكل المنحني يكون على هيئة ناقوس Bell.
2. تتركز القيم او القراءات حول الوسط الحسابي $(\bar{X})$ بحيث تقسمه الى قسمين متساويين.
3. الوسط الحسابي, الوسيط, المنوال, لها نفس القيمة.
4. طرفي المنحني تتناقص بالارتفاع كلما ابتعدنا عن الوسط الحسابي.

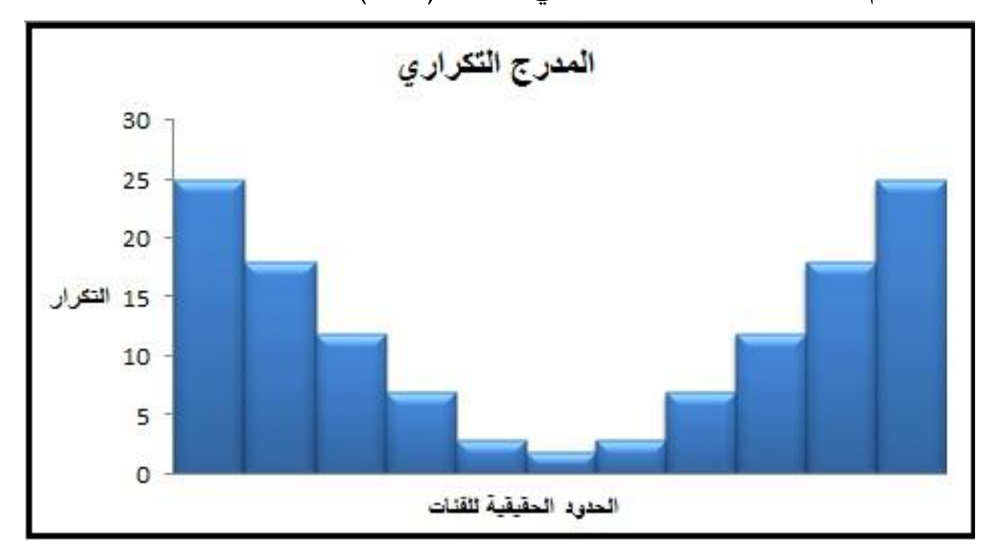
b) منحني على شكل U−Shape :

هو المنحني الذي يتصف بان القيم تتوزع بشكل متماثل على خط المنتصف وعلى يمين ويسار الوسط الحسابي Mean , ولكن نهايات المنحني تمثل او تمتلك قيم عليا , كما في الشكل (3−3):



شكل (3−3) المنحني U−shape

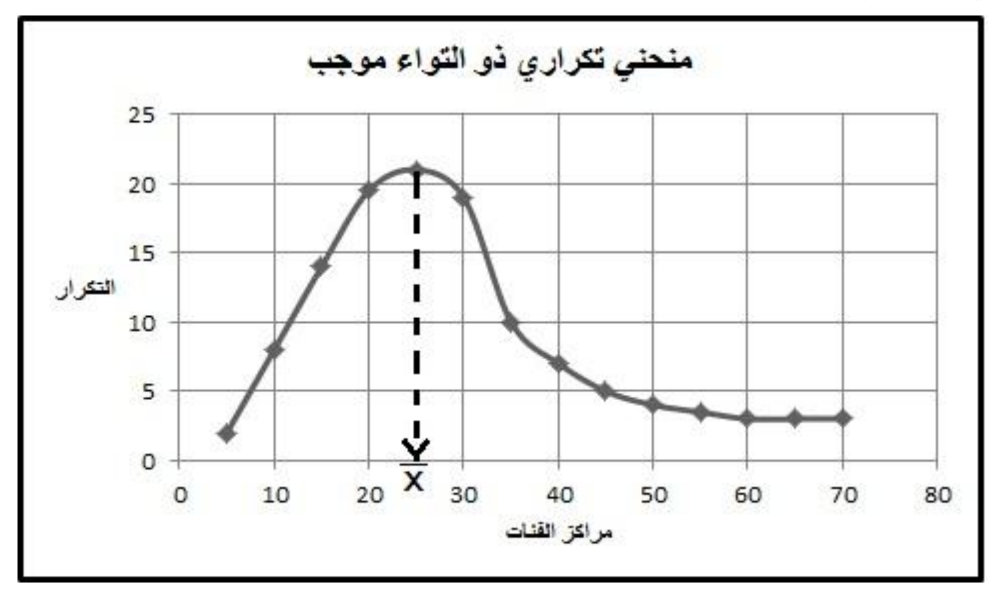
يكون شكل الجدول التكراري الذي يكون منحني يكراري على شكل U–Shape, حيث تمتلك الفئات ذات القيم العالية والفئات ذات القيم القليلة تكرارات عالية كما في الشكل (3–4).



شكل (3–4) شكل المدرج التكراري

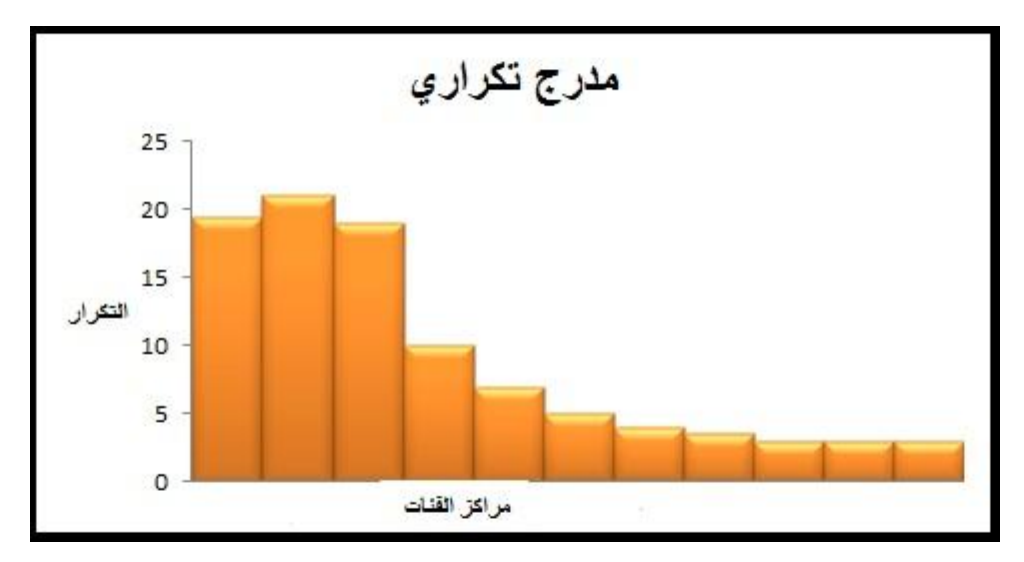
c) منحني ذو التواء موجب Right (Positive) Skewed :

ويسمى كذلك التواء موجب Positive Skewed , وهي المنحنيات التي طرفها الطويل يقع في الجهة اليمنى من المنحني , كما في الشكل (3–5):



شكل (3–5) المنحني ذو التواء موجب

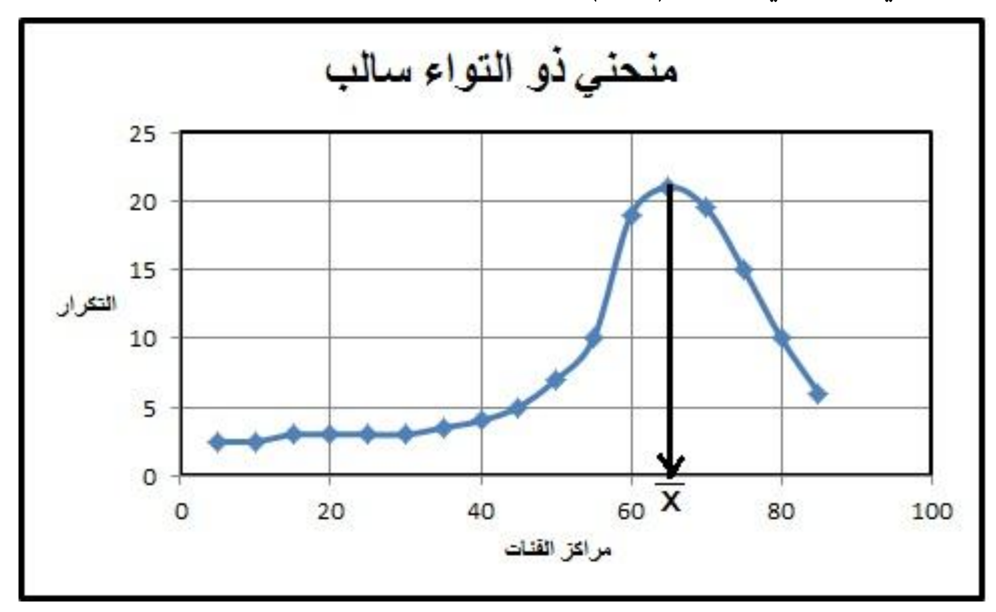
هذا المنحني يعني ان التكرارات القليلة تقع في الفئات او المجاميع ذات القيم العالية , كما في المدرج التكراري (3–6) . هذا النوع من المنحنيات يمتلك قيمة علية واحدة فقط.



شكل (3-6) مدرج تكراري

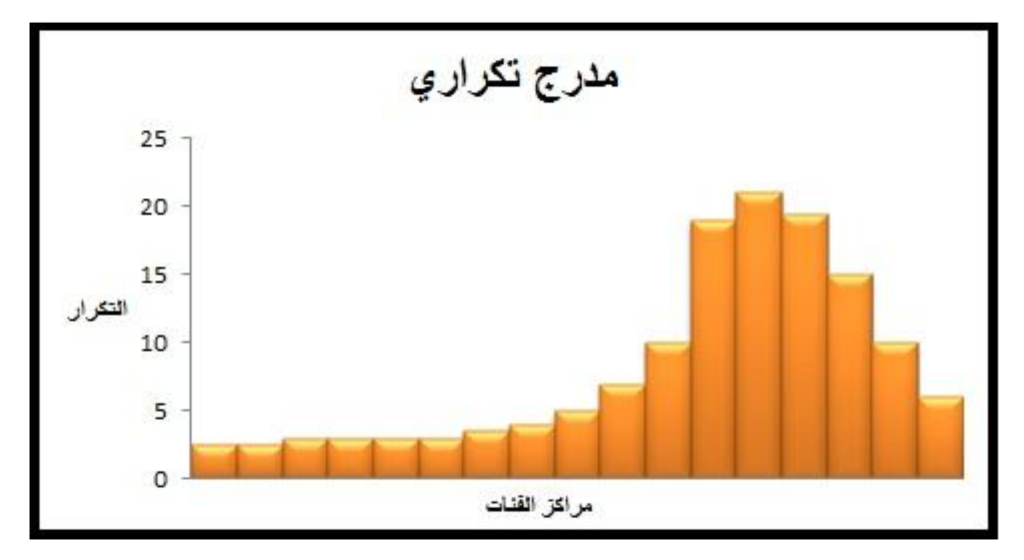
d) منحني ذو التواء سالب Left (Negative) Skewed :

ويسمى كذلك التواء موجب Negative Skewed , وهي المنحنيات التي طرفها الطويل يقع في الجهة اليسرى من المنحني , كما في الشكل (3-7):



شكل (3-7) منحني ذو التواء سالب

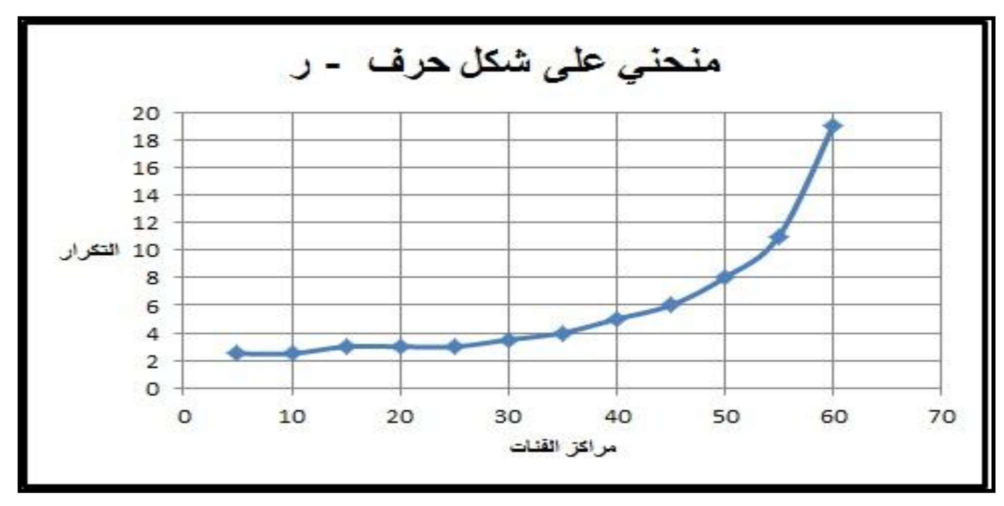
هذا المنحني يعني ان التكرارات القليلة تقع في الفئات او المجاميع ذات القيم الواطئة , كما في المدرج التكراري (3-8) . هذا النوع من المنحنيات يمتلك قيمة علية واحدة فقط.



شكل (3-8) مدرج تكراري

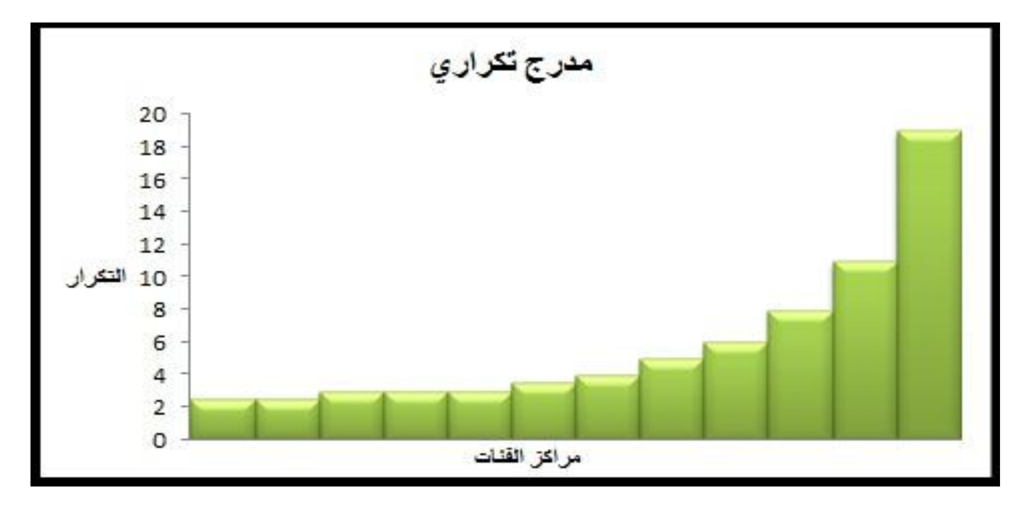
e) <u>منحني على شكل حرف (ر) أو (J-Shape):</u>

هذا المنحني تزداد فيه عدد التكرارات بشكل طردي أو بصورة كبيرة مع زيادة القيم في الفئات الاخيرة او مع زيادة القيم في الحدود العليا للفئات, كما في الشكل (3-9):



شكل (3-9) منحني على شكل حرف (ر)

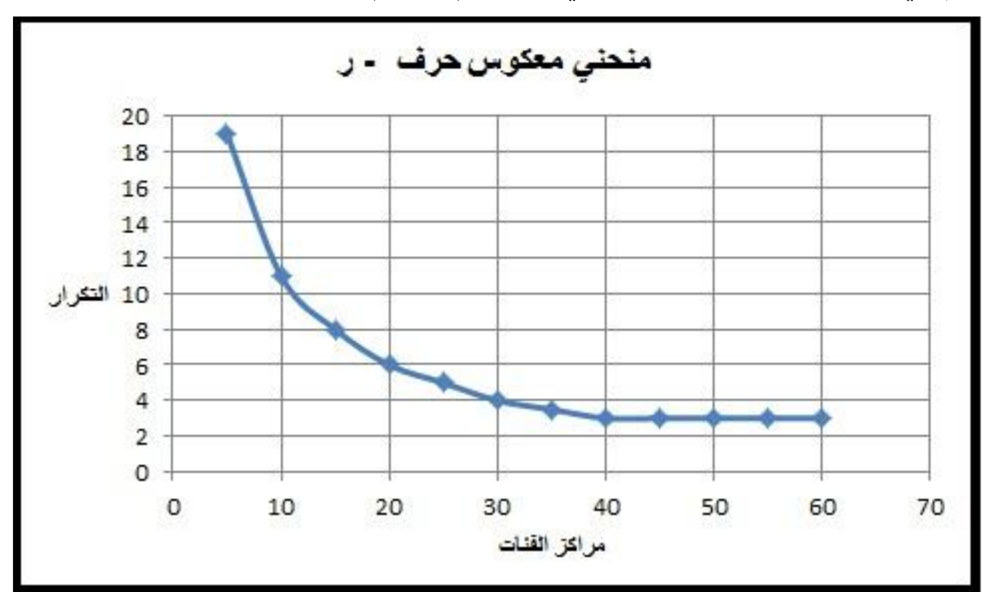
المدرج التكراري (3-10), يوضح كيفية زيادة عدد التكرارات مع زيادة القيم في الفئات الاخيرة ذات القيم العالية.



شكل (3−10) المدرج التكراري

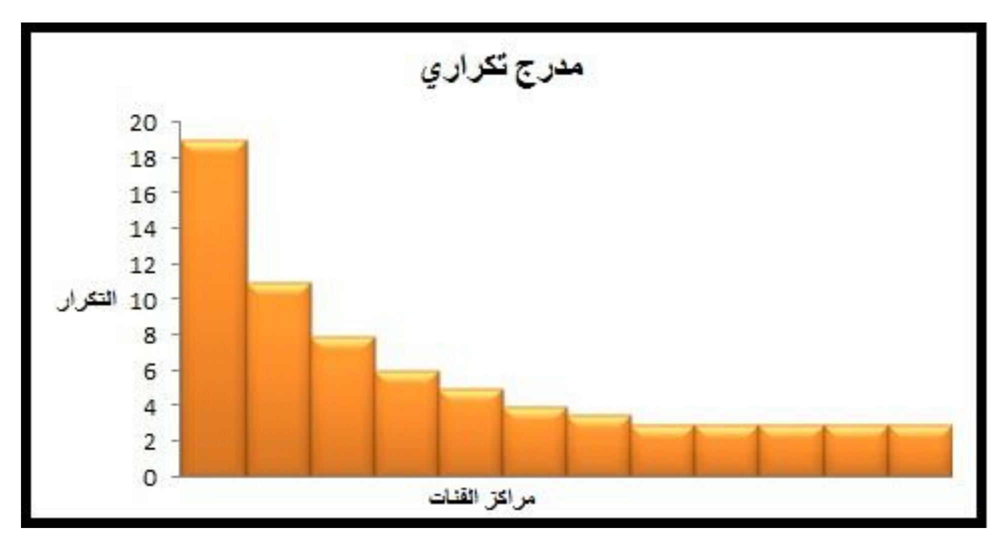
f) <u>منحني على شكل معكوس حرف (ر) او على شكل (Reverse J – Shape) :</u>

هذا المنحني تتناقص فيه عدد التكرارات بشكل طردي أو بصورة كبيرة مع زيادة القيم في الفئات الاخيرة او مع زيادة القيم في الحدود العليا للفئات, كما في الشكل (3−11):



شكل (3−11) منحني على شكل معكوس حرف (ر)

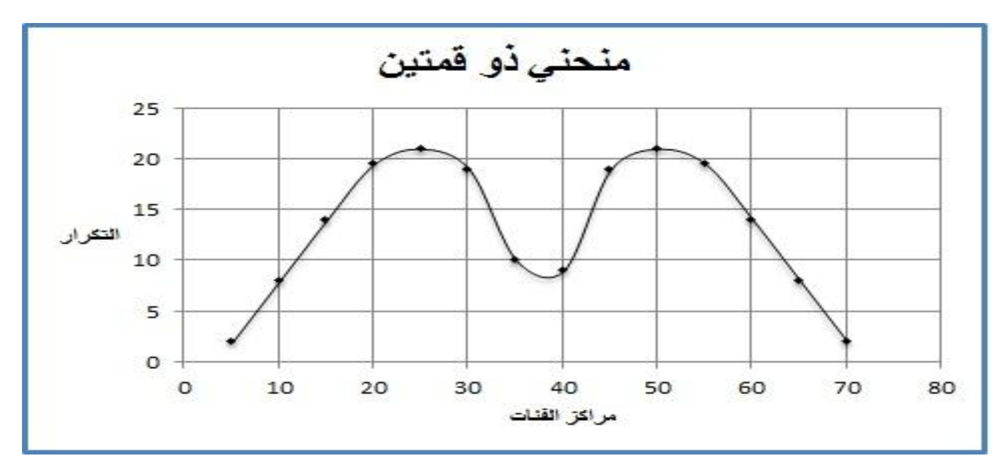
المدرج التكراري (3−12) يوضح كيفية زيادة عدد التكرارات مع زيادة القيم في الفئات الاخيرة ذات القيم العالية.



شكل (3−12) المدرج التكراري

g) <u>منحني ذو قمتين</u> Bimodal Curve:

وهو المنحني الذي يمتلك قمتين او قيمتين عليا, كما في الشكل (3−13):



شكل (3−13) منحني ذو قمتين

المدرج التكراري الذي يكون هذا المنحني يكون بالشكل (3−14):

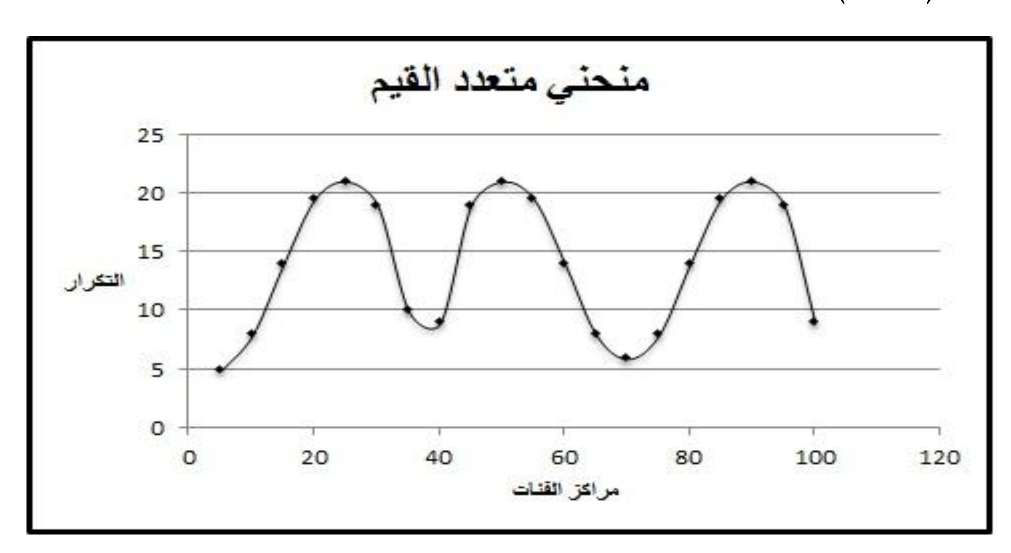


شكل (3−14) المدرج التكراري

h) <u>منحني متعدد القيم</u> Multimodal Curve :

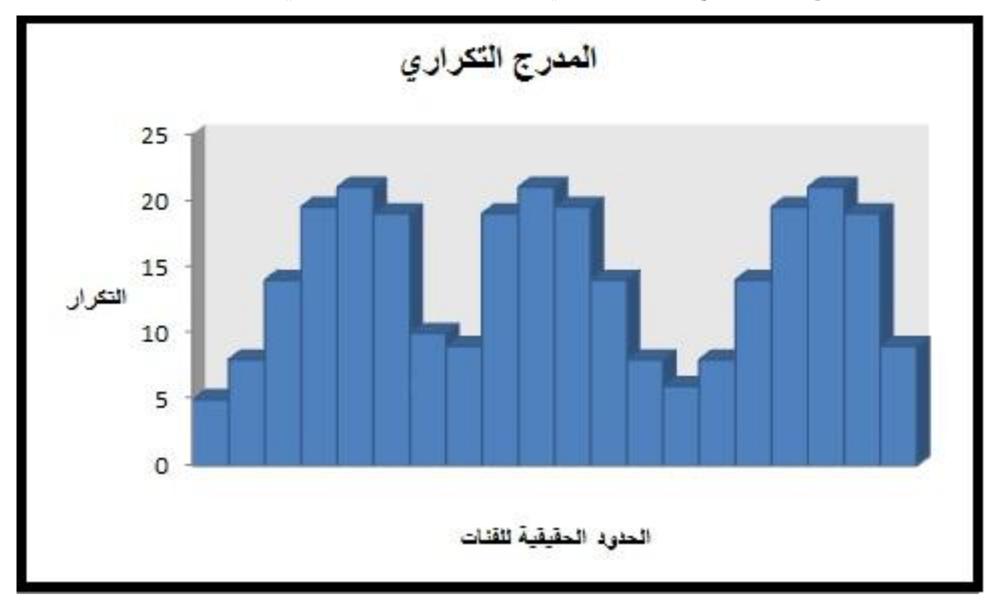
وهو المنحني الذي يمتلك اكثر من قمتين او اكثر من قيمتين عليا, وقد تكون ثلاث, اربع او خمس قمم.....

كما في الشكل (3−15):



شكل (3−15) منحني تكراري متعدد القيم

المدرج التكراري الذي ينتج هذا النوع من المنحني التكراري يكون كما في الشكل (3−16):



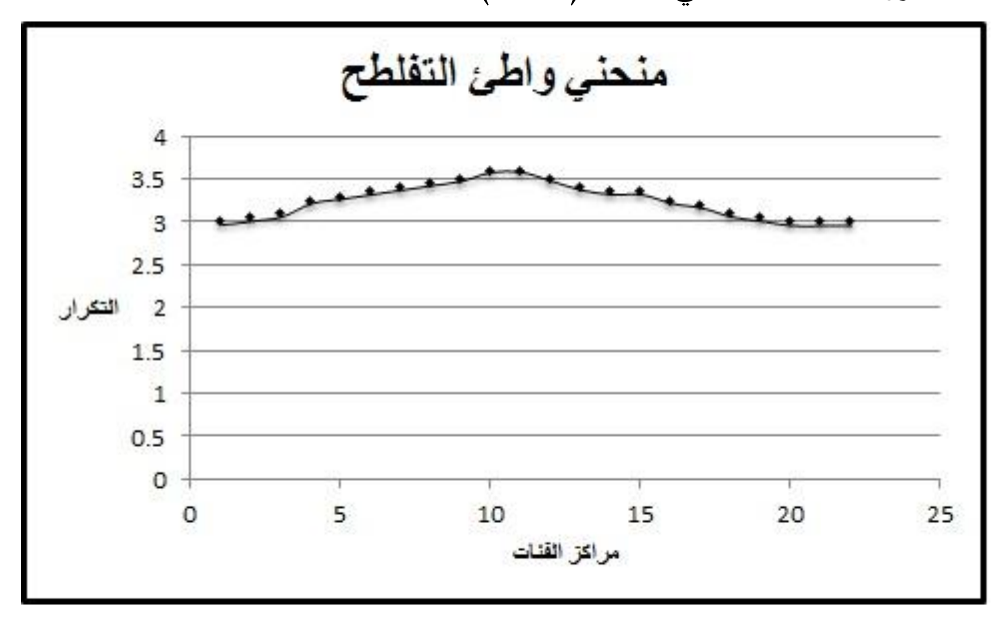
شكل (3−16) المدرج التكراري

i) المنحنيات المتفلطحة Kurtosis Curves :

وهي المنحنيات التي تتميز بامتلاكها انتشار واسع لتكرار القيم على عدد الفئات وذات فئات تكرارية واطئة او متقاربة وتكون على ثلاث انواع:

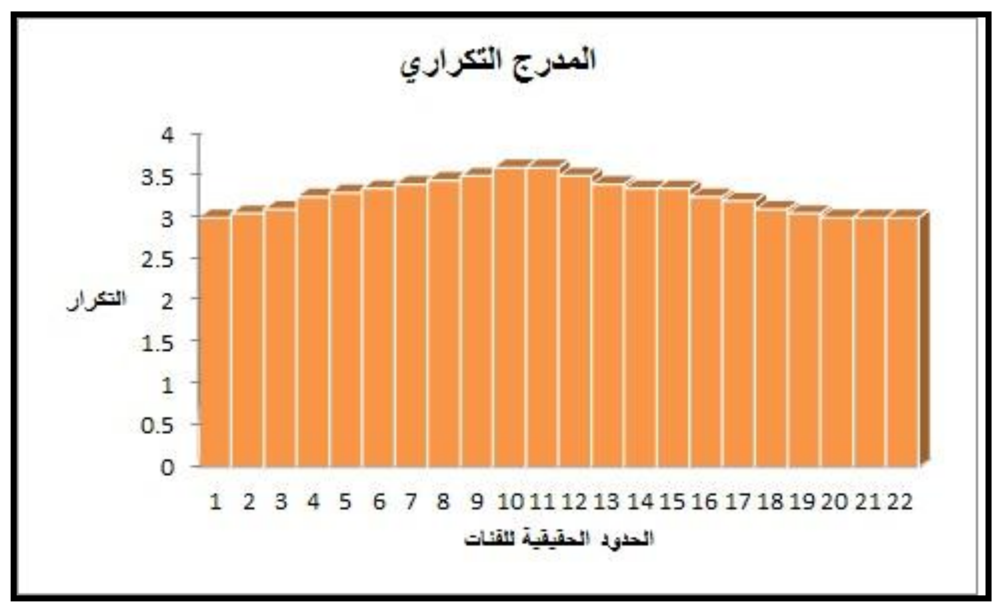
1. منحني واطئ التفلطح Platykurtic:

يتميز هذا المنحني بامتلاكه قمة واطئة او ذات تكرار قليل مع وجود انشار واسع لتكرار القيم على بقية الفئات ذات فروقات قليلة, كما في الشكل (3−17):



شكل (3−17) منحني واطئ التفلطح

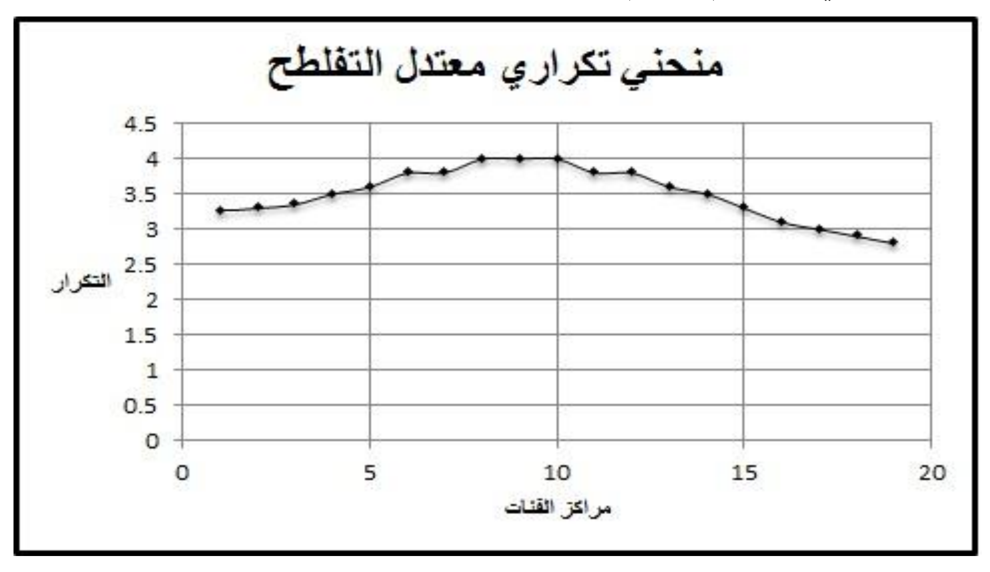
المدرج التكراري الذي يؤدي الى حصول هذا النوع من المنحنيات يكون كما في الشكل (3−18):



شكل (3−18) المدرج التكراري

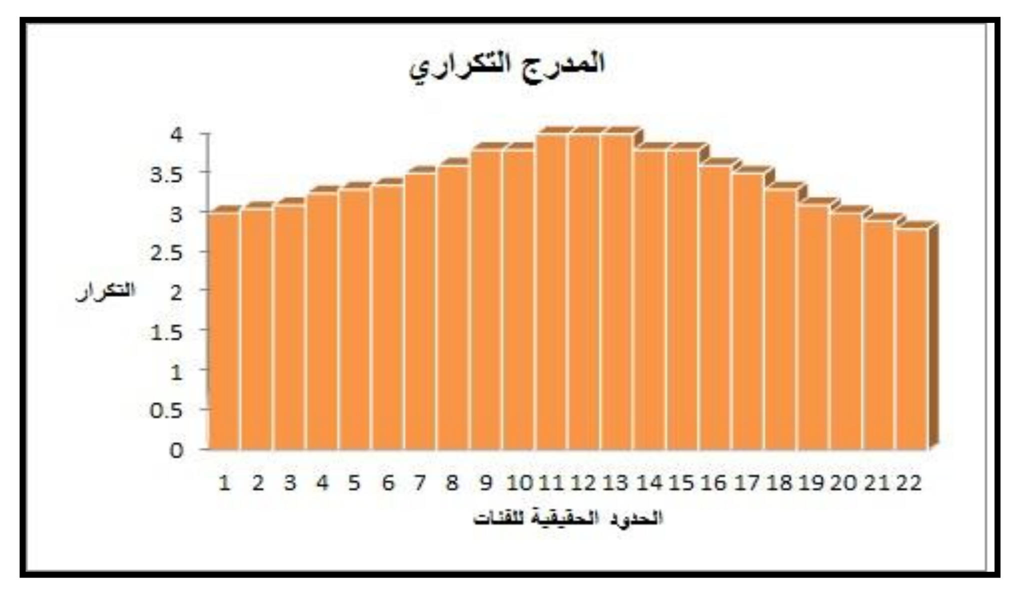
2. منحني تكراري معتدل التفلطح Mesokurtic Curve :

يتميز هذا المنحني بامتلاكه قمة معتدلة او ذات تكرار قليل مع وجود انشار واسع لتكرار القيم على بقية الفئات ذات فروقات متقاربة, كما في الشكل (3−18):



شكل (3−18) منحني معتدل التفلطح

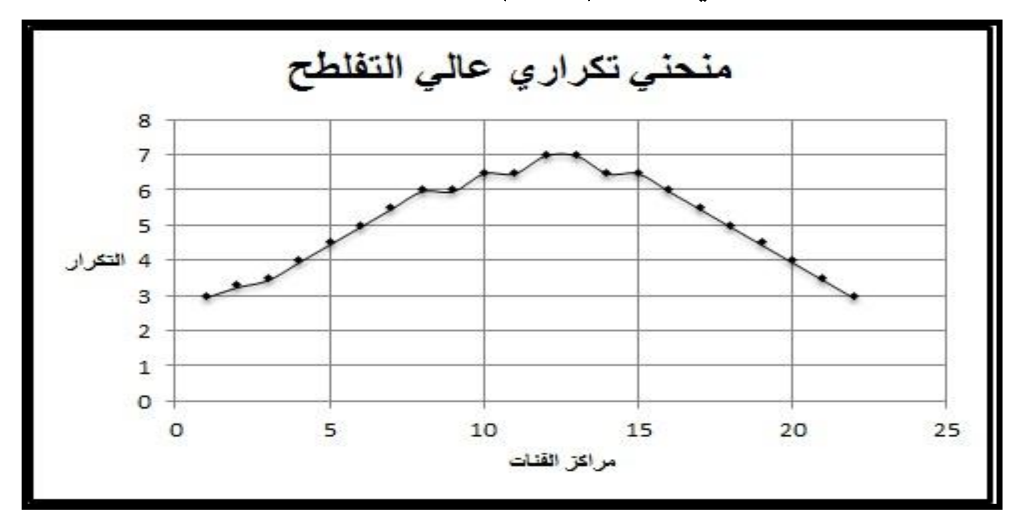
المدرج التكراري الذي ينتج هذا النوع من المنحنيات المعتدلة التفلطح كما في الشكل (3−19):



شكل (3−19) المدرج التكراري

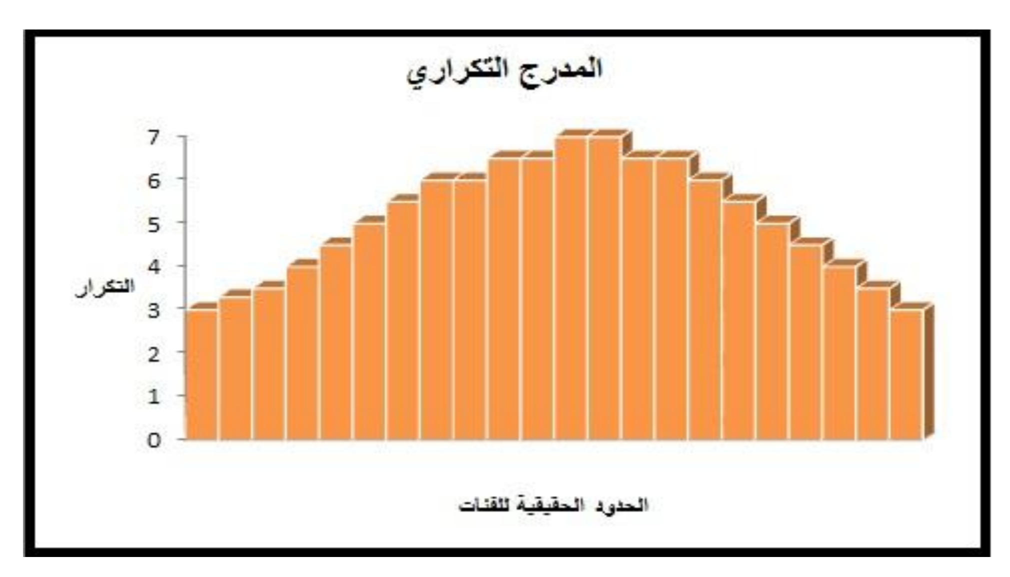
3. منحني عالي التفلطح Leptokurtic Curve :

يتميز هذا المنحني بامتلاكه قمة شبه عالية او ذات تكرار شبه معتدل مع وجود انشار واسع لتكرار القيم على بقية الفئات ذات فروقات شبه متقاربة, كما في الشكل (3−20):



شكل (3−20) منحني عالي التفلطح

المدرج التكراري الذي يكون هذا النوع من المنحني التكراري كما في الشكل (3−21):



شكل (3−21) المنحني التكراري

مسالة تطبيقية:

من الجدول التكراري (3−1), ارسم المنحني التكراري , ثم بين نوع المنحني؟

الفئات	التكرار	الفئات	التكرار
111 - 115	5	161 - 165	12
116 - 120	8	166 - 170	17
121 - 125	14	171 - 175	19
126 - 130	19.5	176 - 180	22
131 - 135	21	181 - 185	22
136 - 140	19.5	186 - 190	20
141 - 145	17	191 - 195	18
146 - 150	12	196 - 200	15
151 - 155	10	201 - 205	10
156 - 160	10	206 – 210	5

الحـــــــــل:

نرسم المنحني التكراري كما في الشكل (3−22) , ثم نبين من الرسم ان المنحني هو من نوع المنحني ذو القمتين Bimodal .



شكل (3−22) منحني ذو القمتين

الفصل الرابع

مقاييس النزعة المركزية أو مقاييس التمركز أو التوسط

Measures of Central Tendency

إن معظم البيانات او النتائج المستحصلة من دراسة ظاهرة معينة عادة ما تميل الى التركيز حول قيمة معينة او تتمركز في الوسط او حول المعدل الحسابي لهذه النتائج او قريبة منه, وفي هذه الحالة يمكن استخدام هذه (القيمة المركزية) لتمثيل هذه المجموعة من البيانات او المقاييس, وبذلك يمكن تعريف مقاييس النزعة المركزية بانها تلك القوانين او العلاقات التي تهتم في ايجاد او حساب او تقدير قيمة معينة تتمركز حولها اغلبية البيانات او النتائج, او بصيغة اخرى هي تلك القوانين التي تهتم في دراسة مقدار ميل النتائج او البيانات الى التمركز او التجمع حول الوسط الحسابي لهذه البيانات.

هذه القيمة هي عبارة عن رقم واحد يمثل المعدل لجميع البيانات المستحصلة من تلك الظاهرة, اذن مهمة الاحصاء هي وصف ظاهرة معينة باستخدام رقم واحد فقط او بمعنى آخر هي اعطاء نتائج موجزة وموثوقة لوصف ظاهرة معينة باستخدام رقم محدد.

بعض المصطلحات تستخدم يوميا بصورة واسعة وشائعة مثل معدل الاسعار , كمية سقوط المطر , معدل الحرارة, معدل تركيز فلز معين , معدل سمك طبقة صخرية , ... الخ وهكذا من هذه المصطلحات ممكن ان نفهم ميل او نزعة القراءات لان تتمركز او تتجمع بالقرب من قيمة معينة , فبدلاً من ان نذكر كافة النتائج ممكن ان نكتفي برقم واحد فقط يمثل جميع هذه القراءات والذي يعطي فكرة واضحة وموثوقة عن طبيعة الظاهرة المدروسة. يجب ان يتوفر في هذه المقاييس الصفات والخصائص التالية لكي يكون المقياس جيداً:

- **A.** ان يعتمد المقياس في حسابه على كل المشاهدات.
- **B.** ان يتوفر فيه الامكانية في التعامل الجبري.
- **C.** ان يكون المقياس سهل الفهم والتعامل.
- **D.** يجب ان لا يتأثر بالقيم الشاذة والمتطرفة.

أهم مقاييس النزعة المركزية هي:-

1. الوسط الحسابي (المتوسط الحسابي) The Arithmetic Mean
2. الوسط الهندسي The Geometric Mean
3. الوسط التوافقي The Harmonic Mean

4. الوسط التربيعي The Quadratic Mean
5. الوسيط The Median
6. المنوال The Mode

كل مقياس من هذه المقييس يفضل استخدامه في حالات معينة او محددة ولا يفضل استخدامه في حالت اخرى.

1. **<u>الوسط الحسابي (المتوسط الحسابي) The Arithmetic Mean:</u>**

هو القيمة التي تمثل معدل القيم او النتائج والتي تنتج من حاصل قسمة مجموع تلك القيم على عددها او هو حاصل قسمة مجموع قيم مفردات العينة على حجم العينة, يرمز الى الوسط الحسابي بالرمز ($\overline{X}$), والصيغة الرياضية له في حالة البيانات غير المبوبة هي:

$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{\sum X}{n} = \frac{X1 + X2 + X3 + \cdots + Xn}{n}$$

حيث ان Xi تمثل القيم او القراءات المستحصلة و n تمثل عدد تلك القيم.

مثال على ذلك اذا كانت لدينا القيم التاليــة: 84 , 91 , 72 , 68 , 87 , 78

أوجد الوسط الحسابي لها.

$$\overline{X} = \frac{\sum Xi}{n} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = \frac{480}{6} = 80$$

في حالة الجداول التكرارية يتم استخراج الوسط الحسابي بموجب الخطوات التالية:-

1. تعيين مراكز الفئات (X_i) ,حيث يتم اضافة عمود الى الجدول التكراري يمثل مركز الفئة .
2. ضرب مركز كل فئة في تكرار تلك الفئة = ($fi \; x \; Xi$).
3. قسمة مجموع (حاصل ضرب مركز كل فئة في مقدار تكرارها) على مجموع التكرارات $\bar{X} = \left[\frac{\sum Fi \; x \; Xi}{\sum fi}\right]$
   لكي نحصل على الوسط الحسابي للبيانات في الجدول التكراري.

مثال:

من الجدول التكراري رقم (4−1) : أوجـد الوسط الحسابي.

جدول رقم (4−1)

الفئات Classes	التكرار fi	مركز الفئة Xi	fi *x* Xi
31 – 40	1	35.5	35.5
41 – 50	2	45.5	91.0
51 – 60	5	55.5	277.5
61 – 70	15	65.5	282.5
71 - 80	25	75.5	1887.5
81 – 90	20	85.5	1710.0
91 – 100	12	95.5	1146.0
المجموع ∑	∑ = 80		∑ = 6130

$$\therefore \ \overline{X} = \frac{\sum(fi\ x\ Xi)}{\sum fi} = \frac{6130}{80} = 76.62$$

من خصائص الوسط الحسابي ان مجموع انحرافات القيم عن وسطها الحسابي تساوي صفر.

$$\sum\left[Xi - \overline{X}\right] = 0$$

اما في الجداول التكرارية فتكون :

$$\sum fi\ x\left[Xi - \overline{X}\right] = 0$$

2. **الوسط الهندسي The Geometric Mean :**

يعرف الوسط الهندسي الذي يرمز له بالرمز (G) لمجموعة من القيم التي يرمز لها بالرمز (N) بانه الجذر مرفوع بعدد القيم لحاصل ضرب كافة القيم او القراءات, كما في الصيغة الرياضية التالية:

$$G = \sqrt[N]{X1\ x\ X2\ x\ X3\ x \ldots\ldots x\ Xn}$$

مثال على ذلك ان الوسط الهندسي للقيم (8 , 4 , 2) سوف يكون:

$$G = \sqrt[3]{2\ x\ 4\ x\ 6} = \sqrt[3]{64} = 4$$

عادة في التطبيق العملي يتم حساب الوسط الهندسي باستخدام اللوغاريتمات , مثال على ذلك:

$$G = \sqrt[7]{3\ x\ 5\ x\ 6\ x\ 6\ x\ 7\ x\ 10\ 12} = \sqrt[7]{453600}$$

$$or \quad \log G = \frac{1}{7}\ \log 453600 = \frac{1}{7}\,(5.6567) = 0.8081$$

$$\therefore G = 6.43$$

في حين ان الوسط الحسابي لهذه القيم هو:

$$\overline{X} = \frac{\sum Xi}{n} = \frac{3\ x\ 5\ x\ 6\ x\ 6\ x\ 7\ x\ 10\ x\ 12}{7} = 7$$

<u>من خصائص الوسط الهندسي :</u>

A. تكون قيمة الوسط الهندسي موجبة دائماً.

B. اصغر قيمة من الوسط الحسابي.

C. لا يمكن ايجاد قيمة الوسط الهندسي الا اذا كانت مجموع القيم موجبة.

D. يستخدم الوسط الهندسي في ايجاد الارقام القياسية للاسعار وايجاد معدل التغيير في المبيعات او السكان.

3. **<u>الوسط التوافقي The Harmonic Mean :</u>**

يعرف الوسط التوافقي بانه مقلوب الوسط الحسابي لمقلوب القيم او النتائج, ويرمز له بالرمز ($\overline{H}$), والصيغة الرياضية للوسط التوافقي هي:

$$\overline{H} = \frac{N}{\sum \frac{1}{\overline{\overline{X}}}}$$

من خصائص الوسط التوافقي هو ان يكون اقل من الوسط الحسابي, ويستخدم عادة في ايجاد معدل سعر سعر سلعة معينة عندما يكون هناك تفاوت كبير في اسعار تلك السلعة.

<u>مثال على ذلك:</u>

اشترى مزارع بذور القمح بسعر اجمالي قدره (100000) دينار من كل من الشركات التالية:

الشركة الاولى كان سعر الطن الواحد من بذور القمح = 20000 دينار

الشركة الثانية كان سعر الطن الواحد من بذور القمح = 25000 دينار

الشركة الثالثة كان سعر الطن الواحد من بذور القمح = 50000 دينار

ما هو سعر متوسط سعر الطن الواحد من بذور القمح؟

الحـــــــل:

نستخدم الوسط التوافقي في ايجاد متوسط سعر الطن الواحد في حالة وجود تباين كبير في الاسعار:

$$\bar{H} = \frac{N}{\sum \frac{1}{\bar{X}}} = \frac{3}{\frac{1}{20000} + \frac{1}{25000} + \frac{1}{50000}} = \frac{3}{1.1x10^{-4}} = 27272.7 \text{ Dinar}$$

الوسط التوافقي في حالة الجداول التكرارية :

يمكن ايجاد الوسط التوافقي في حالة الجداول التكرارية من العلاقة الرياضية التالية:

$$\bar{H} = \frac{\text{مجموع التكرارات}}{\sum \left[\frac{\text{التكرار}}{\text{مركز الفئة}}\right]} = \frac{\sum fi}{\sum \left[\frac{fi}{Xi}\right]}$$

مثـــــال :

من الجدول التكراري (4−2), اوجد الوسط التوافقي.

جدول (4−2)

Classes	Frequency fi	Mid-Point Xi
60 – 62	5	61
63 – 65	18	64
66 – 68	42	67
69 – 71	27	70
72 - 74	8	73

الحـــــــل:

$$\bar{H} = \frac{\text{مجموع التكرارات}}{\sum \left[\frac{\text{التكرار}}{\text{مركز الفئة}}\right]} = \frac{\sum fi}{\sum \left[\frac{fi}{Xi}\right]} = \frac{100}{\frac{5}{61} + \frac{18}{64} + \frac{42}{67} + \frac{27}{70} + \frac{8}{73}} = \frac{100}{1.4855} = 67.3$$

4. الوسط التربيعي The Quadratic Mean :

يعرف الوسط التربيعي بانه الجذر التربيعي للوسط الحسابي لمربعات القيم Root Mean Square .

ويرمز له بالرمز (.R.M.S). العلاقة الرياضية الخاصة به هي :

$$R.M.S. = \sqrt{\frac{\sum(\bar{X})^2}{N}}$$

مثال:

أوجد الوسط التربيعي للقيم التالية: (3 ,5 ,6 ,6 ,7 ,10 ,12)

$$R.M.S. = \sqrt{\frac{\sum(\bar{X})^2}{N}} = \sqrt{\frac{(3)^2+(5)^2+(6)^2+(6)^2+(7)^2+(10)^2+(12)^2}{7}}$$

$$= \sqrt{57} = 7.55$$

الوسط التربيعي في حالة الجداول التكرارية:

لغرض ايجاد الوسط التربيعي في حالة الجداول التكرارية نتبع الحل كما في العلاقة الرياضية التالية:

$$R.M.S. = \sqrt{\frac{\sum(Xi)^2 x\, fi}{\sum fi}} = \sqrt{\frac{\sum \text{مربع مركز الفئة } x \text{ التكرار}}{\sum \text{التكرار}}}$$

مثال:

من الجدول التكراري (4−3) ,اوجد الوسط التربيعي.

Classes	Frequency fi	Mid-Point Xi
60 – 62	5	61
63 – 65	18	64
66 – 68	42	67
69 – 71	27	70
72 - 74	8	73

الحـــــــل:

يتم تنظيم جدول تكراري كما في الجدول (4−4) , لغرض استخراج المتغيرات المطلوبة لتعويضها في المعادلة الرياضية:

جدول (4−4)

Classes	Frequency fi	Mid-Point Xi	(Xi)²	fi x (Xi)²
60 – 62	5	61	3721	18605
63 – 65	18	64	4096	73728
66 – 68	42	67	4489	188538
69 – 71	27	70	4900	132300
72 - 74	8	73	5329	42632
	∑ 100			∑455803

$$.M.S.=\sqrt{\frac{\sum(Xi)^2 x\, fi}{\sum fi}}=\sqrt{\frac{\sum \text{التكرار } x \text{ مربع مركز الفئة}}{\sum \text{التكرار}}}=\sqrt{\frac{455803}{100}}=67.51$$

5. **الوســـيط The Median :**

يعرف الوسيط بانه القيمة الواقعة في وسط مجموعة من القيم او القراءات عندما يتم ترتيبها تصاعديا او تنازلياً. ويرمز له بالرمز (Me) , وهو القيمة التي تقسم مفردات العينة الى قسمين متساويين بعد ترتيبها تصاعدياً أو تنازلياً بحيث يكون عدد المفردات الاصغر منها في القيمة مساويا لعدد المفردات الاكبر منها في القيمة.

في حالة الاعداد الصحيحة : اذا كان عدد القراءات فردي بهذه الحالة تكون قيمة الوسيط هي القيمة التي تقع في وسط القراءات.

مثال على الاعداد الصحيحة : إذا كان لدينا عدد فردي من القيم (حجم العينة فردي) وتم ترتيبها تصاعدياً, كما يلي:

(2 , 3 , 4 , 5 , 6 , 10 , 15)

ترتيب قيمة الوسيط يستخرج من العلاقة التالية: $Me=\frac{n+1}{2}$ حيث ان n هي عدد القيم.

$$\therefore\ Me=\frac{n+1}{2}=\frac{7+1}{2}=\frac{8}{2}=4$$

اذن ترتيب قيمة الوسيط هي القيمة الرابعة والتي تساوي (5).

إذا كان لدينا عدد زوجي من القيم (حجم العينة زوجي), وتم ترتيبها تصاعديا او تنازليا وكما يلي:

(1 , 3 , 6 , 8 , 11 , 15 , 17 , 22)

بهذه الحالة يتم استخراج ترتيب قيمة الوسيط والتي تمثل معدل القيمتين الوسطيتين باستخدام العلاقة التالية:

ترتيب القيمة الاولى = $\left[\frac{n}{2}\right]$, ترتيب القيمة الثانية = $\left[\frac{n}{2}+1\right]$

نستخرج ترتيب القيمة الاولى = $\frac{n}{2}$ = $\frac{8}{2}$ = 4 , إذن القيمة التي ترتيبها (4) تساوي (8).

نستخرج ترتيب القيمة الثانية =$\left[\frac{n}{2}+1\right]$ = $\frac{8}{2}$ + 1= 1+4 = 5, إذن القيمة التي تريبها(5)تساوي (11)

∴ قيمة السيط بهذه الحالة تساوي $\frac{8+11}{2}$ = $\frac{19}{2}$ = 9.5

في حالة الجداول التكرارية أو البيانات المبوبة, يتم ايجاد قيمة الوسيط من الصيغة الرياضية التالية:

$$Me = L1 + \left(\frac{\frac{N}{2} - (\sum f)}{fmedian}\right) x\, C$$

L_1 = الحد الادنى الحقيقي لفئة الوسيط

N = مجموع التكرارات

$\sum f$ = محموع التكرارات لكافة الفئات لغاية بداية فئة الوسيط

$f\ median$ = تكرار فئة الوسيط

C = طول فئة الوسيط

<u>كيفية ايجاد قيمة الوسيط من جدول التوزيع التكراري:</u>

1. عمل جدول تكراري
2. ايجاد ترتيب الوسيط وهو =$\frac{\sum f}{2}$ = $\frac{\text{مجموع التكرارات}}{2}$
3. ايجاد التكرار التجمعي الصاعد
4. نحدد فئة الوسيط (وهي الفئة التي يقع ترتيب الوسيط بين حديها) ويستحسن ان نأخذ الحدود الحقيقية لها.
5. نستخرج قيم المتغيرات المطلوبة لغرض تعويضها في القانون , ثم نستخرج قيمة الوسيط.

مثال :أوجد قيمة الوسيط من الجدول التكراري (4−5):

جدول (4−5)

الفئات Classes	التكرار *fi*	تكرار تجمعي صاعد
60 – 62	5	5
63 – 65	18	23
66 – 68	42	65
69 – 71	27	92
72 – 74	8	100
المجموع	∑ 100	

<u>الحــل:</u>

A. نستخرج ترتيب الوسيط = $\frac{\sum fi}{2} = \frac{100}{2} = 50$, هذا يعني ان قيمة الوسيط تقع عند الترتيب (50) , ومن التكرار التجمعي الصاعد (الذي تم ترتيب القيم بصورة تصاعدية), نجد ان الترتيب (50) يقع ضمن الفئة الثالثة , اي بعد التكرار التجمعي الذي هو (23) لغاية التكرار التجمعي (65) اذ ان الترتيب (50) يقع ضمن الفئة الثالثة (68 – 66), بهذه تم تحديد فئة الوسيط.

B. بعد تحديد فئة الوسيط يتم استخراج بقية المتغيرات والتي هي:

$L1 = 65.5$ *الحد الادنى الحقيقي لفئة الوسيط*

$\sum f = 23$ مجموع التكرارات لغاية بداية فئة الوسيط

$f\ median = 42$ *تكرار* فئة الوسيط

$C = 3$ طول فئة الوسيط

$N = 100$ مجموع التكرارات

C. نطبق ونعوض في القانون:

$$Me = L1 + \left(\frac{\frac{N}{2} - (\sum f)}{fmedian}\right) x\, C = 65.5 + \left(\frac{\frac{100}{2} - (23)}{42}\right) x\, 3 = 67.4$$

الطريقة الثانية في ايجاد قيمة الوسيط هي طريقة الرسم او طريقة التمثيل الصوري للبيانات, وهي طريقة مهمة وسريعة يتم من خلالها استخراج قيمة الوسيط مباشرة , حيث يتم رسم المنحني التكراري التجمعي الصاعد

والمنحني التكراري التجمعي النازل في نفس الرسم على الورقة البيانية وعند إنزال عمود من نقطة تقاطع هذين المنحنيين على المحور السيني (مراكز الفئات), فان نقطة التقاطع هذه تمثل قيمة الوسيط.

مثال: من الجدول التكراري (4-6) اوجد قيمة الوسيط بطريقة الرسم البياني.

جدول رقم (4-6)

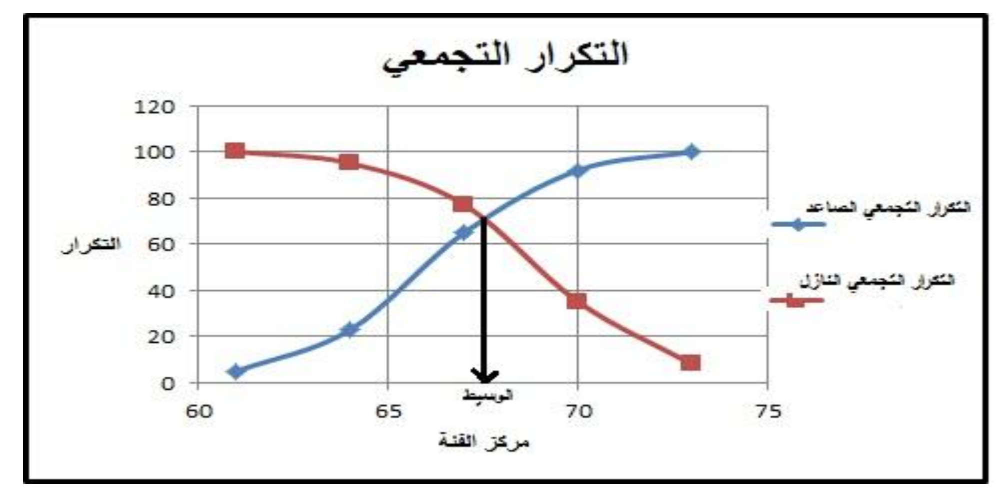
الفئات Classes	التكرار *fi*
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
المجموع	∑ 100

الحل: نستخرج قيمة التكرار التجمعي الصاعد والنازل كما في الجدول (4-7), ثم نرسم المنحني التكراري

جدول (4-7) التكرار التجمعي الصاعد والنازل

الفئات Classes	التكرار *fi*	تكرار تجمعي صاعد	تكرار تجمعي نازل
60 – 62	5	5	100
63 – 65	18	23	95
66 – 68	42	65	77
69 – 71	27	92	35
72 – 74	8	100	8
المجموع	∑ 100		

التجمعي الصاعد والنازل في مرتسم واحد كما في الشكل (4-1) , حيث يكون العمود النازل من نقطة تقاطع المنحنيين على المحور السيني , نقطة الالتقاء تمثل قيمة الوسيط. قيمة الوسيط تساوي 67.5 .



شكل (4-1) المنحني التكراري التجمعي الصاعد والنازل

مسالة تطبيقية :

اذا كان لدينا الجدول التكراري (4−8):

جدول (4−8)

الفئات	التكرار
118 – 126	3
127 – 135	5
136 – 144	9
145 – 153	12
154 – 162	5
163 – 171	4
172 - 180	2

اوجد قيمة الوسيط بطريقتين:

a. بالطريقة الرياضية باستخدام قانون الوسيط

b. بطريقة الرسم باستخدام المنحني التكراري التجمعي

الحـــــــل:

a. اعتماداً على الجدول التكراري(4−8) , نوجد التكرار التجمعي الصاعد والنازل

b. نوجد فئة الوسيط

c. نوجد قيم المتغيرات ثم نطبق قانون الوسيط

d. يتم عمل جدول تكراري يتضمن القيم التكرارية التجمعية الصاعدة والنازلة , كما في الجدول (4−9).

جدول (4−9) ايجاد التكرار التجمعي الصاعد والنازل

الفئات	التكرار	مركز الفئة	التكرار التجمعي الصاعد	التكرار التجمعي النازل
118 – 126	3	122	3	40
127 – 135	5	131	8	37
136 – 144	9	140	17	32
145 – 153	12	149	29	23
154 – 162	5	158	34	11
163 – 171	4	167	38	6
172 - 180	2	176	40	2

نستخرج ترتيب الوسيط = $\frac{\sum fi}{2} = \frac{40}{2} = 20$, إذن ترتيب الوسيط يقع في الفئة الرابعة (153 − 145),

L1 = 144.5 *الحد الادنى الحقيقي لفئة الوسيط*

$\sum f$ = 17 مجموع التكرارات لغاية بداية فئة الوسيط

$f\ median$ = 12 *تكرار* فئة الوسيط

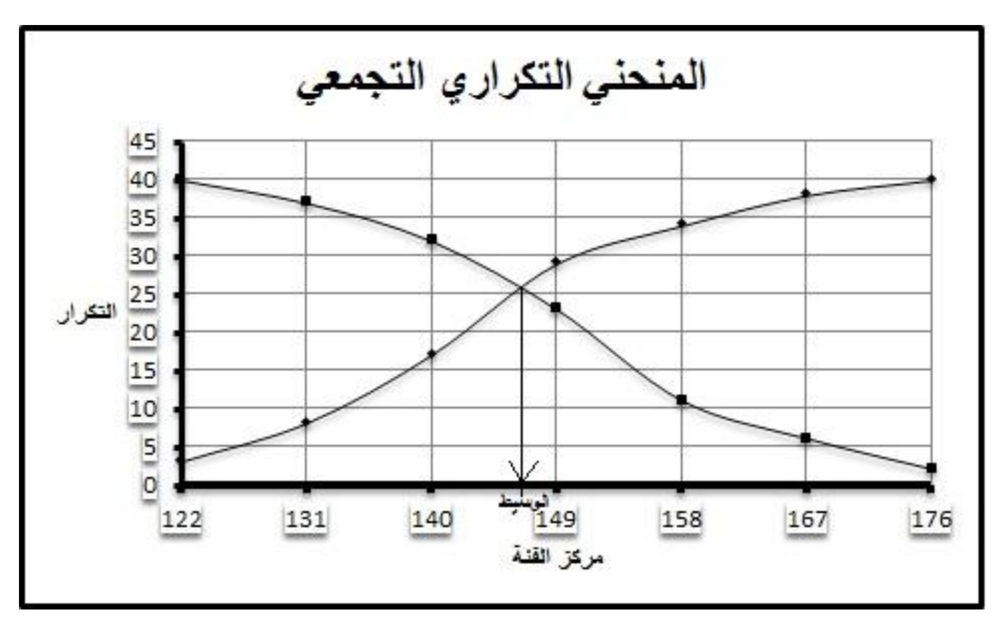
C = 9 طول فئة الوسيط

N = 40 مجموع التكرارات

$$Me = L1 + \left(\frac{\frac{N}{2} - (\sum f)}{fmedian}\right) x\, C \;=\; 144.5 + \left(\frac{\frac{40}{2} - (17)}{12}\right) x\, 9 = 146.75$$

قيمة الوسيط تساوي 146.75

نرسم المنحني التكراري التجمعي الصاعد والنازل في مرتسم واحد, شكل (4−2), ثم نستخرج قيمة الوسيط من نقطة تقاطع المنحني التكراري التراكمي ومسقطها على المحور السيني والتي تساوي تقريبا (147) .



شكل (4−2) مرتسم المنحني التكراري التجمعي الصاعد والنازل

<u>مثال اثرائي:</u>

أوجد قيمة الوسيط لاجور العمال اليومية في احدى الشركات بالدينار وحسب الجدول التكراري (4−10).

جدول (4−10) توزيع الاجور اليومية للعمال

الاجور (دينار)	عدد العمال (التكرار)
10 − 19	3
20 − 29	6
30 − 39	10
40 − 49	15
50 − 59	8
60 − 69	5
70 − 79	3
المجموع	50

<u>الجواب</u> : Me = 44 دينار

6. **<u>المنــــــــــــــــوال The Mode:</u>**

يعرف المنوال بانه القيمة الاكثر تكرارا في مجموعة من القيم او القراءات, ويرمز له بالرمز *(Mo)* . اذا كانت لدينا مجموعة من القراءات او البيانات غير المبوبة, قد لا يوجد فيها منوال او قد يوجد فيها قيمة واحدة للمنوال تسمى (Unimodal) , او قد توجد قيمتين للمنوال (Bimodal) , أو قد يكون لهذه القيم اكثر من منوالين تسمى (Multimodal) .

<u>مثـــال (1)</u>: أوجد قيمة المنوال لمجموعة من القيم التالية:

(2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18)

المنوال في هذه الحالة هي القيمة الاكثر تكرارا وهو الرقم (9).

<u>مثـــال (2)</u>: أوجد المنوال لمجموعة من القيم التالية:

(2. 3, 4, 4, 4, 5, 5, 7, 7, 7, 9)

المنوال في هذه الحالة هي القيمة (4) والقيمة (7) , اي يوجد فيها منوالين وتسمى (Bimodal).

في حالة الجداول التكرارية او البيانات المبوبة, يمكن استخراج المنوال وفق الصيغة الرياضية التالية:

$$\text{المنوال} = \boldsymbol{Mo} = \boldsymbol{L1} + \left(\frac{\boldsymbol{\Delta 1}}{\boldsymbol{\Delta 1} + \boldsymbol{\Delta 2}}\right) \boldsymbol{x}\, \boldsymbol{C}$$

L_1 = الحد الادنى الحقيقي لفئة المنوال

$\boldsymbol{\Delta 1}$ = الفرق بين تكرار فئة المنوال والفئة السابقة لها

$\boldsymbol{\Delta 2}$ = الفرق بين تكرار فئة المنوال والفئة اللاحقة لها

C = طول الفئة

<u>مثـــال (3)</u>:

من الجدول التكراري (4–10) , اوجد قيمة المنوال.

الفئات Classes	التكرار fi
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
المجموع	∑ 100

الحــــــــل:

نلاحظ ان الفئة الثالثة (68 – 66) هي الفئة التي تمتلك اكثر التكرارات , وهي بذلك تكون فئة المنوال.

ثم نستخرج بقية المتغيرات لغرض تطبيق القانون وكما يلي:

L_1 = 65.5 =الحد الادنى الحقيقي لفئة المنوال

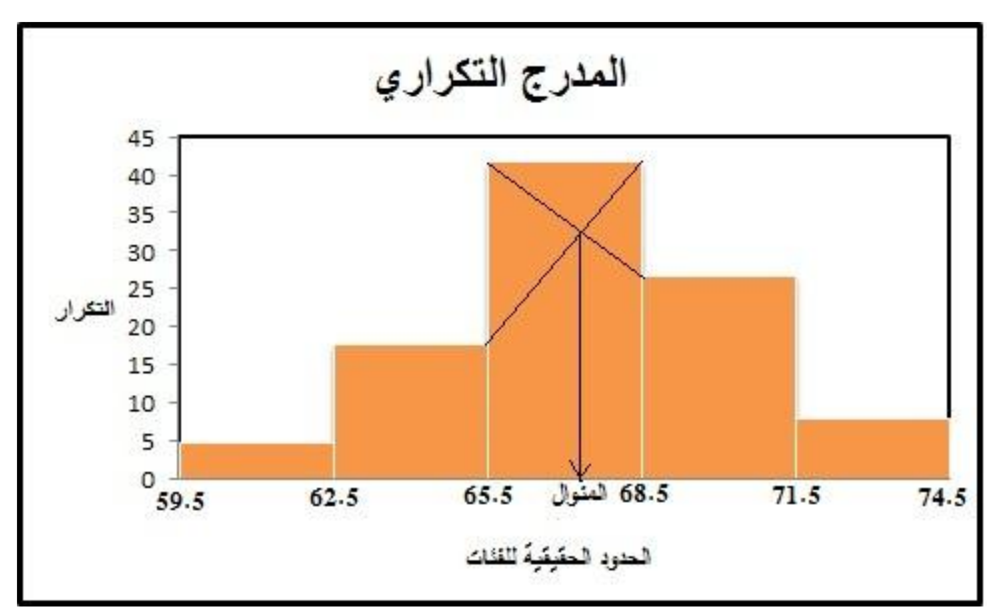
$\Delta 1$ = 42−18=24 = الفرق بين تكرار فئة المنوال والفئة السابقة لها

$\Delta 2$ = 42−27=15 = الفرق بين تكرار فئة المنوال والفئة اللاحقة لها

C = 3= طول الفئة

$$\text{المنوال} = Mo = L1 + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) x\, C = 65.5 + \left(\frac{24}{24 + 15}\right) x\, 3 = 67.3$$

الطريقة الثانية التي تستخدم في ايجاد قيمة المنوال هي باستخدام المدرج التكراري, حيث ان المستطيل الاعلى هو الذي يمثل فئة المنوال. باستخدام هذه الطريقة يتم انزال خط مستقيم من الزوايا العليا لمستطيل الفئة المنوالية لتقطع الزاوية العليا المقابلة لها من المستطيلات المتجاورة , أو بصيغة اخرى نصل رؤوس المستطيلات المجاورة لمستطيل الفئة المنوالية ببعضها فتتقاطع في نقطة ضمن مستطيل الفئة المنوالية , ثم نسقط منها عمود على المحور الافقي الذي يمثل الحدود الحقيقية للفئات او مراكز الفئات متكون هي قيمة المنوال وكما في الشكل (3−4):



شكل (3−4) ايجاد قيمة المنوال

قيمة المنوال بهذه الحالة تساوي تقريبا 67.3 التي يمكن قراءتها من نقطة تقاطع العمود النازل على المحور السيني.

مسالة تطبيقية (1):

الجدول التكراري (4−11), يمثل توزيع رواتب الموظفين في احدى الشركات الى (65) موظف. أوجد:

a. الوسيط بطريقتين مختلفتين.

b. المنوال بطريقتين مختلفتين.

c. الوسط الحسابي.

جدول (4−11)

الفئـات	التكرار
50 – 59	8
60 – 69	10
79 – 70	16
80 – 89	14
90 – 99	10
100 – 109	5
110 – 119	2
المجموع	65

الحـــــل:

a. نستخرج المتغيرات المطلوبة من الجدول التكراري (4−11), ويتم تسجيلها كما في الجدول التكراري (4−12).

جدول تكراري (4−12)

الفئـات	التكرار fi	مركز الفئة Xi	التكرار التجمعي الصاعد	التكرار التجمعي النازل	$fi \times Xi$
50 – 59	8	54.5	8	65	436
60 – 69	10	64.5	18	57	645
70 - 79	16	74.5	34	47	1192
80 – 89	14	84.5	48	31	1183
90 – 99	10	94.5	58	17	945
100 – 109	5	104.5	63	7	5225
110 – 119	2	114.5	65	2	229
المجموع	$\sum 65$				$\sum$ **9855**

نوجد الوسيط من المعادلة الرياضية الخاصة به:

$$Me = L1 + \left(\frac{\frac{N}{2} - (\sum f)}{fmedian}\right) x\ C \ = \ 69.5 + \left(\frac{\frac{65}{2} - (18)}{16}\right) x\ 10 = 78.5$$

ترتيب فئة الوسيط = $\frac{65}{2} = 32.5$, اذن فئة الوسيط هي (79 − 70). الحدود الحقيقية لها هي (69.5 −79.5)

الحد الادنى الحقيقي لفئة الوسيط = L1 = 69.5

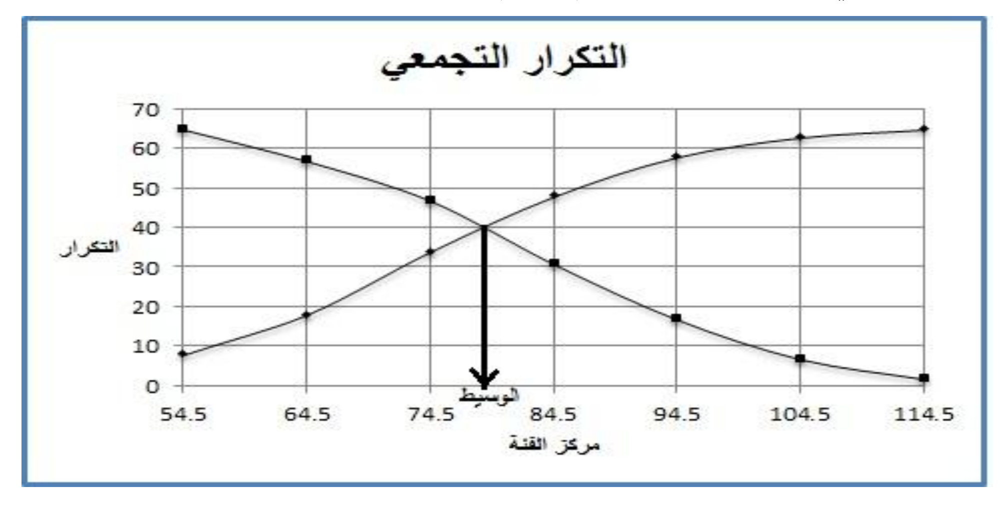
التكرار التجمعي عند بداية فئة الوسيط = $\sum f$ = 18

تكرار فئة الوسيط = $fmedian$ = 16

طول الفئة = C = 10

مجموع التكرارات = N = 65

نرسم التكرار التجمعي الصاعد والنازل, كما في الشكل (4−4), ومن نقطة تقاطع المنحنيين نرسم العمود النازل وتقاطعه مع المحور السيني نقرأ قيمة الوسيط = (78.5) .



شكل (4−4) استخراج قيمة الوسيط

b. المنوال = $Mo = L1 + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) x\ C \ = \ 69.5 + \left(\frac{6}{6+2}\right) x\ 10 = 77.0$

الفئة المنوالية هي الفئة الاكثر تكرارا وهي (79 − 70)

الحد الادنى الحقيقي لفئة المنوال = L1 = 69.5

الفرق بين تكرار فئة المنوال والفئة السابقة لها = 16 − 10 = 6 = $\Delta 1$

الفرق بين تكرار فئة المنوال والفئة اللاحقة لها = 16 − 14 = 2 = $\Delta 2$

طول الفئة = C = 10

لغرض ايجاد المنوال من خلال رسم المدرج التكراري كما في الشكل (4-5), وكما يلي:



شكل (4-5) ايجاد قيمة المنوال

المنوال = 77

<u>مسالة تطبيقية (2):</u>

من الجدول التكراري (4-13), أوجد:

1. الوسط الحسابي ($\bar{X}$)
2. الوسط التوافقي ($\bar{H}$)
3. الوسط التربيعي (R.M.S.)
4. الوسيط (Me)
5. المنوال (Mo)

جدول تكراري (4-13)

التكرار	الفئات
3	40 – 45
9	46 – 51
20	52 – 57
35	58 – 63
15	64 – 69
8	70 – 75
90	المجموع

<u>الحـــــــل :</u> نستخرج المتغيرات المطلوبة وتدرج كما في الجدول التكراري (4−14), لغرض تعويضها في القوانين الخاصة بها وكما يلي:

جدول تكراري (4−14)

الفئات	التكرار fi	مركز الفئة Xi	تكرار تجمعي صاعد	تكرار تجمعي نازل	fi x Xi	$\frac{fi}{Xi}$	Xi² x fi
40 – 45	3	42.5	3	90	127.5	0.070	5418.75
46 – 51	9	48.5	12	87	436.5	0.185	21170.2
52 – 57	20	54.5	32	78	1090.0	0.367	59405.0
58 – 63	35	60.5	67	58	2117.5	0.578	128108.7
64 – 69	15	66.5	82	23	997.5	0.225	66333.7
70 – 75	8	72.5	90	8	580.5	0.110	42050.0
المجموع	90				5349.0	1.535	322486.3

1. $\bar{X} = \frac{\sum fi\ x\ Xi}{N} = \frac{5349.0}{90} = 59.5$

2. $\bar{H} = \frac{\sum fi}{\sum\frac{fi}{Xi}} = \frac{90}{1.535} = 58.6$

3. $R.M.S. = \sqrt{\frac{\sum(Xi)^2\ x\ fi}{\sum fi}} = \sqrt{\frac{322486.3}{90}} = \sqrt{3583.18} = 59.8$

4. الوسيط $= Me = L1 + \left(\frac{\frac{N}{2}-(\sum f)}{fmedian}\right) x\ C = 57.5 + \left(\frac{\frac{90}{2}-(32)}{35}\right) x\ 6 = 59.7$

ترتيب فئة الوسيط = $\frac{90}{2}$ = **45**, اذن فئة الوسيط هي (63 − 58). الحدود الحقيقية لها هي (57.5 −63.5)

الحد الادنى الحقيقي لفئة الوسيط = 57.5 = L1

التكرار التجمعي عند بداية فئة الوسيط = 32 = $\sum f$

تكرار فئة الوسيط = 35 = $fmedian$

طول الفئة = 6 = C

مجموع التكرارات = 90 = N

5. المنوال = $Mo = L1 + \left(\frac{\Delta1}{\Delta1+\Delta2}\right) x\ C = 57.5 + \left(\frac{15}{15+20}\right) x\ 6 = 60.07$

الحد الادنى الحقيقي لفئة المنوال = 57.5 = L1

الفرق بين تكرار فئة المنوال والفئة السابقة لها = 20−35 = 15 = $\Delta1$

الفرق بين تكرار فئة المنوال والفئة اللاحقة لها = 15−35 = 20 = $\Delta2$

طول الفئة = 6 = C

الفصل الخامس

مقاييس التفلطح والالتواء

Measures of Skewness and Kurtosis

تعرف مقاييس التفلطح والالتواء بانها تلك القوانين التي تحدد شكل المنحني التكراري من حيث التماثل او الالتواء والانحراف الى احدى الجهات (Skewness), وكذلك تحديد تدبب القمة للمنحني او تفلطحها (Kurtosis), بمعنى آخر هو بعد المنحني التكراري عن التماثل. ويقصد بالتماثل انه اذا اسقطنا عموداً من قمة المنحني التكراري وتم تقسيمه الى قسمين متماثلين متطابقين فان التوزيع يكون متماثلاً, وبالعكس فان التوزيع يكون غير متماثل إذغ كان المنحني ملتوي الى احدى الجهات (يميناً أو يساراً).

هناك مصطلح يستخدم في حساب هذه المقاييس وهو ما يسمى بالعزوم (Moments) التي تعتبر احدى مقاييس الالتواء.

العزوم Moments

اذا كان لدينا n من المشاهدات او النماذج التابعة للمتغير X , فان :

هي متغيرات من دراسة ظاهرة معينة عدد نماذجها = n , X1, X2, X3,……, Xn

1- فان العزم الرائي حول الصفر في حالة البيانات غير المبوبة هو: $\overline{X^r} = \frac{\sum Xi^r}{n}$

العزم الاول حول الصفر هو نفسه الوسط الحسابي , اي ان:

$$\bar{X} = \frac{\sum Xi}{n} = \text{الوسط الحسابي}$$

اما العزم الثاني حول الصفر يساوي: $\overline{X^2} = \frac{\sum Xi^2}{n}$, وهكذا.

اما في حالة البيانات المبوبة , اذا كانت (X1, X2, X3, …..) تمثل مراكز الفئات في جدول التوزيع التكراري , وان تكرارات هذه الفئات هي (f1, f2, f3, ……) فان العزم الرائي حول الصفر هو:

$$\overline{X^r} = \frac{\sum fi \ x \ Xi^r}{\sum fi}$$

2- في حالة العزم الرائي حول الوسط الحسابي The rth moment about the mean في حالة البيانات غير المبوبة:

تكون الصيغة الرياضية له :

$$m_r = \frac{\sum (Xi - \bar{X})^r}{n}$$

العزم الرائي الاول (اذا كانت r=1), بهذه الحالة يساوي صفر لان حاصل جمع فروقات القيم عن الوسط الحسابي يساوي صفر كما في المعادلة التالية:

$$m_1 = \frac{\sum(Xi - \overline{X})}{n} = 0$$

اما اذا كانت (r=2) اي العزم الرائي الثاني حول الوسط الحسابي فان العزم الرائي يساوي التباين كما في المعادلة التالية:

$$m_2 = \frac{\sum(Xi - \overline{X})^2}{n} = \text{The Variance}$$

وهكذا بالنسبة لبقية العزوم , والتي تسمى بالعزم الرائي حول الوسط الحسابي.

في حالة البيانات المبوبة فان العزم الرائي حول الوسط الحسابي سوف يكون:

$$m_r = \frac{\sum fi\,(Xi - \overline{X})^r}{\sum fi}$$

<u>مثـــال(1) :</u>

اوجد العزم الرائي الاول, الثاني, الثالث, الرابع حول الصفر للاعداد التالية: 2, 3, 7, 8, 10

العزم الرائي الاول = الوسط الحسابي = $\overline{X} = \frac{\sum Xi}{n} = \frac{2+3+7+8+10}{5} = \frac{30}{5} = 6$

العزم الرائي الثاني = $\overline{X^2} = \frac{\sum Xi^2}{n} = \frac{2^2+3^2+7^2+8^2+10^2}{5} = \frac{226}{5} = 45.2$

العزم الرائي الثالث = $\overline{X^3} = \frac{\sum Xi^3}{n} = \frac{2^3+3^3+7^3+8^3+10^3}{5} = \frac{1890}{5} = 378$

العزم الرائي الرابع = $\overline{X^4} = \frac{\sum Xi^4}{n} = \frac{2^4+3^4+7^4+8^4+10^4}{5} = \frac{16594}{5} = 3318.8$

مثـــال(1) :

اوجد العزم الرائي الاول, الثاني, الثالث, الرابع حول الوسط الحسابي للاعداد التالية: 2, 3, 7, 8, 10

الوسط الحسابي للاعداد اعلاه = 6

$$m_1 = \frac{\sum(Xi - \bar{X})}{n} = \frac{(2-6)+(3-6)+(7-6)+(8-6)+(10-6)}{5} = \frac{0}{5}$$

$= 0 = m_1 =$ يساوي صفر دائما

$$m_2 = \frac{\sum(Xi - \bar{X})^2}{n} = \frac{(2-6)^2+(3-6)^2+(7-6)^2+(8-6)^2+(10-6)^2}{5}$$

$= \frac{46}{5} = 9.2 = m_2 =$ يساوي التباين دائما

$$m_3 = \frac{\sum(Xi - \bar{X})^3}{n} = \frac{(2-6)^3+(3-6)^3+(7-6)^3+(8-6)^3+(10-6)^3}{5}$$

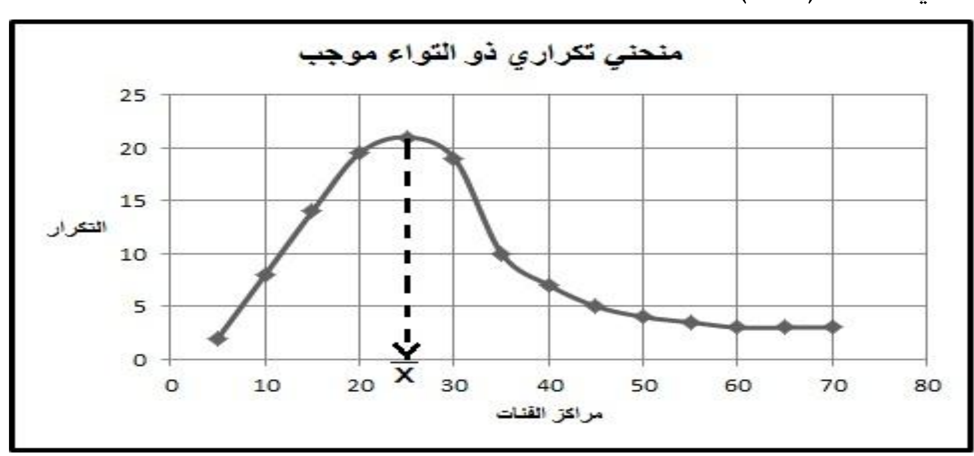
$$= \frac{-18}{5} = -3.5$$

$$m_4 = \frac{\sum(Xi - \bar{X})^4}{n} = \frac{(2-6)^4+(3-6)^4+(7-6)^4+(8-6)^4+(10-6)^4}{5}$$

$$= \frac{610}{5} = 122$$

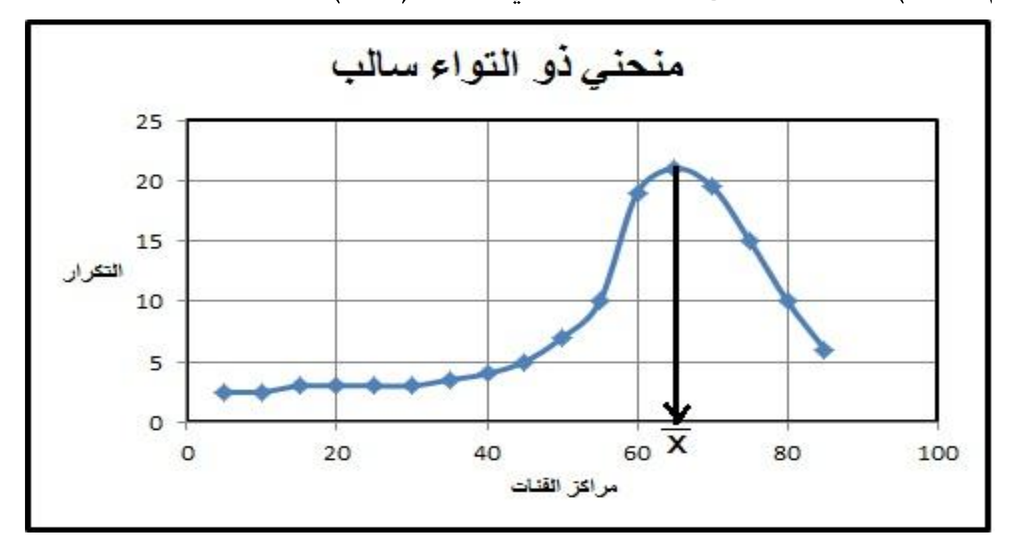
مقاييس الالتواء Measures of Skewness

يعرف الالتواء بانه انحراف منحني التوزيع التكراري عن التماثل او عن المنحني الطبيعي, وقد يكون الالتواء موجب (اي الالتواء الى اليمين) أو الالتواء سالب (اي الالتواء الى اليسار). منحني الالتواء ذو الالتواء الموجب تكون معظم القيم او القراءات متمركزة في الجهة اليسرى (في الفئات الدنيا ذات القيم القليلة) وطرفه يمتد الى اليمين , كما في الشكل (5−1):



شكل (5−1) منحني ذو التواء موجب

أما منحني التوزيع التكراري ذو الالتواء السالب فان معظم القيم او القراءات متمركزة في الجهة اليمنى (في الفئات العليا ذات القيم العالية) وطرفه يمتد الى اليسار , كما في الشكل (5−2):



شكل (5−2) منحني ذو التواء سالب

<u>اهم مقاييس الالتواء هي:</u>

1) باستخدام المنوال , ويرمز له بالرمز $(\boldsymbol{\alpha_1})$, ويسمى معامل الالتواء الاول.

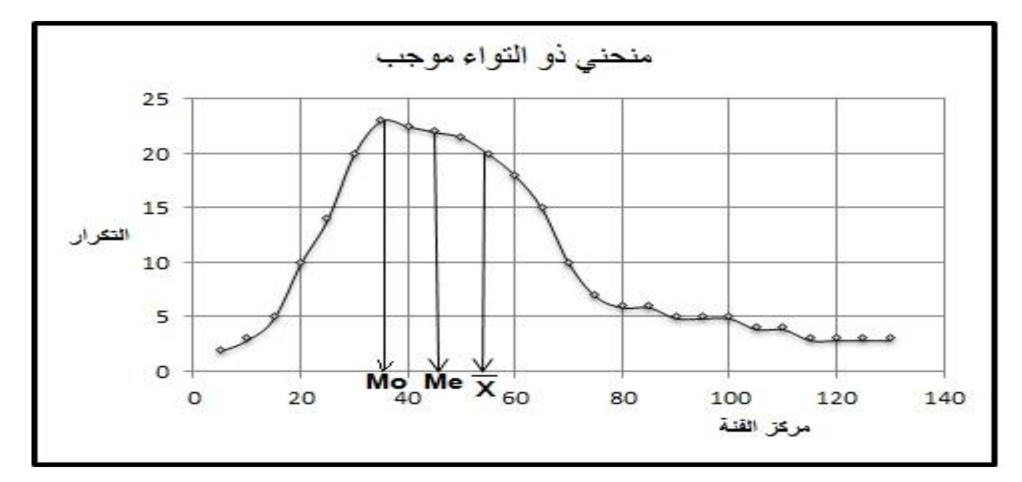
2) باستخدام الوسيط , ويرمز له بالرمز $(\boldsymbol{\alpha_2})$, ويسمى معامل الالتواء الثاني.

3) باستخدام العزوم , ويرمز له بالرمز $(\boldsymbol{\alpha_3})$, ويسمى معامل الالتواء الثالث.

علما بانه في جميع هذه الطرق الثلاث يكون الالتواء موجب عندما يكون معامل الالتواء موجب, ويكون الالتواء سالب عندما يكون معامل الالتواء سالب ومتماثلا عندما يكون معامل الالتواء صفر.

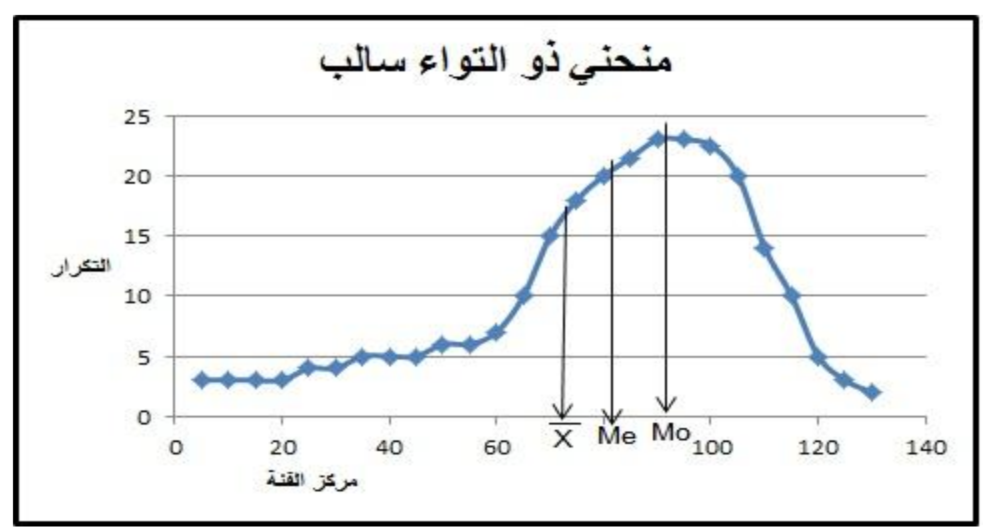
معامل الالتواء الاول = $\frac{\text{الوسط الحسابي} - \text{المنوال}}{\text{الانحراف القياسي}} = \frac{(\bar{X} - Mo)}{\sigma} = \boldsymbol{\alpha_1}$

من خلال تطبيق هذا القانون نلاحظ اذا كانت قيمة الوسط الحسابي اكبر من قيمة المنوال , كانت النتيجة موجبة وبالتالي كان الالتواء موجب, كما في الشكل (5−3):



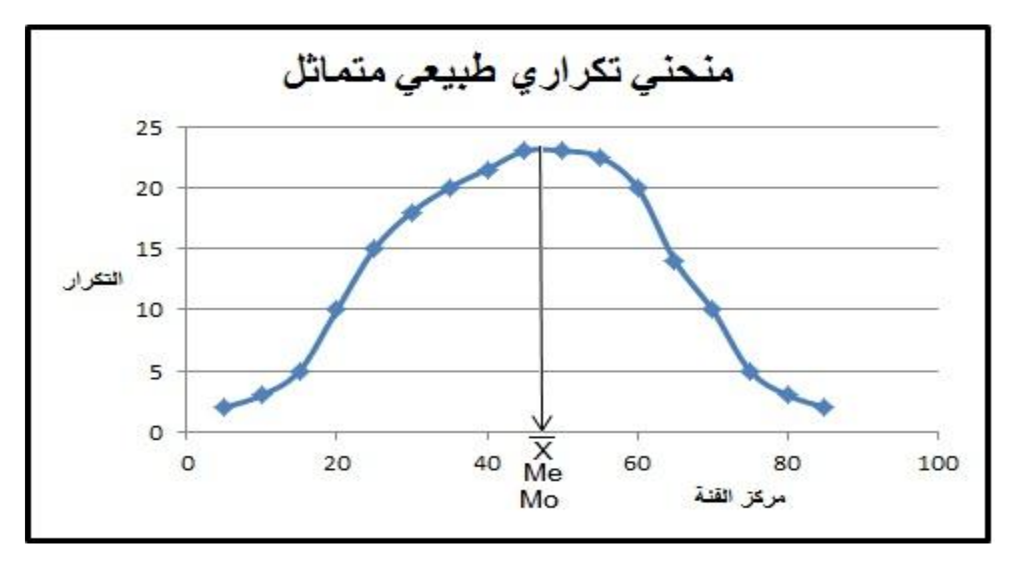
شكل (5-3) منحني ذو الالتواء الموجب يبين العلاقة بين الوسط الحسابي والوسيط والمنوال

اما اذا كانت قيمة الوسط الحسابي اقل من المنوال كانت النتيجة سالبة وبالتالي كان الالتواء سالب, كما في الشكل (5-4):



شكل (5-4) منحني ذو الالتواء السالب يبين العلاقة بين الوسط الحسابي والوسيط والمنوال

اما اذا تساوت قيمة الوسط الحسابي والمنوال والوسيط نحصل في هذه الحالة على منحني تكراري متماثل, كما في الشكل (5-5):



شكل (5-5) المنحني المتماثل يبين العلاقة بين الوسط الحسابي والوسيط والمنوال

معامل الالتواء الثاني = $\frac{3\,x\,(\text{الوسط الحسابي} - \text{الوسيط})}{\text{الانحراف القياسي}} = \frac{3\,x\,(\bar{X} - Me)}{\sigma} = \alpha_2$

في هذه الحالة اذا كانت قيمة الوسط الحسابي اكبر من الوسيط , كانت نتيجة معامل الالتواء الثاني موجبة وبالتالي كان الالتواء موجب, واذا كانت قيمة الوسط الحسابي اقل من الوسيط كان الالتواء سالب. اما اذا كانت قيمة الوسط الحسابي مساوية الى قيمة الوسيط فان معامل الالتواء الثاني يساوي صفر وبالتالي يكون التوزيع متماثلاً.

معامل الالتواء الثالث باستخدام العزوم Moment = $\frac{\text{العزم الثالث حول الوسط الحسابي}}{\text{مكعب الجذر التربيعي للعزم الثاني حول الوسط الحسابي}}$

$$\alpha_3 = \frac{m_3}{\sqrt{m_2^3}} =$$

فاذا كانت قيمة $\alpha_{3=0}$ فذلك يدل على وجود تماثل في توزيع القراءات وبالتالي فان المنحني يكون طبيعي , انا اذا كانت قيمة m_3 موجبة فالتوزيع التكراري يكون ملتوي التواءا موجبا, اما اذا كانت m_3 سالبة فان التوزيع سوف يكون ذو التواء سالب بهذه الحالة.

هناك قوانين تستخدم لحساب معامل الالتواء ومن اهمها:

$$Skewness = \frac{\sum(Xi - \bar{X})^3}{n\sigma^3}$$

ملاحظة:

1. عادة ما نحصل على نتائج مختلفة لمعاملات الالتواء , وهذا لا يناقض بعضه إذ ان كل معامل يقيس الالتواء على اساس يخالف المعاملات الاخرى.
2. عند مقارنة التواء توزيعات مختلفة يجب استخدام نفس معامل الالتواء عند المقارنة.

مثـــال (1):

احسب معامل الالتواء الاول والثاني والثالث في جدول التوزيع التكراري (5−1):

جدول (5−1)

الفئات Classes	التكرار *fi*
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
المجموع	∑ 100

الحـــــل :

الوسط الحسابي = $\bar{X}$ = 67.45

الوسيط = Me = 67.40

المنوال = M0 = 67.30

الانحراف القياسي = σ = 2.92

معامل الالتواء الاول = α_1 =
$$\alpha_1 = \frac{(\bar{X}-Mo)}{\sigma} = \frac{(67.45-67.30)}{2.92} = \frac{0.15}{2.92} = 0.051$$

معمل الالتواء الثاني = α_2 =
$$\alpha_2 = \frac{3\,x\,(\bar{X}-Me)}{\sigma} = \frac{3\,x(67.45-67.40)}{2.92} = \frac{0.15}{2.92} = 0.05$$

التباين = $m_2 = \frac{\sum(Xi-\bar{X})^2}{n} = 8.5275$

$$m_3 = \frac{\sum(Xi-\bar{X})^3}{n} = -2.6932$$

معامل الالتواء الثالث = α_3 =
$$\alpha_3 = \frac{m_3}{\sqrt{m_2^3}} = \frac{-2.6932}{\sqrt{8.5275^3}} = -0.14$$

من هذا نستنتج ان المنحني ملتوي التواء سالب (الى اليسار) وذلك لان (α_3) سالبة.

مثال (2):

أوجد معامل الالتواء الاول والثاني لاجور العمال المدرجة في الجدول التكراري التالي:

الفئات (اجور العمال)	التكرار (عدد العمال)
10 – 19	3
20 – 29	6
30 – 39	10
40 – 49	15
50 – 59	8
60 – 69	5
70 – 79	3
المجموع	50

الحـــــــــل:

الوسط الحسابي = $\bar{X}$ = 44.20

الوسيط = Me = 44.17

المنوال = M0 = 44.0

الانحراف القياسي = σ = 15.21

معامل الالتواء الاول = α_1 =

$$\alpha_1 = \frac{(\bar{X}-Mo)}{\sigma} = \frac{(44.20-44.17)}{15.21} = \frac{0.03}{15.21} = 0.002$$

معمل الالتواء الثاني = α_2 =

$$\alpha_2 = \frac{3(\bar{X}-Me)}{\sigma} = \frac{3(44.2-44.0)}{15.21} = \frac{0.6}{2.92} = 0.04$$

مثال (3):

الجدول التكراري التالي يمثل التوزيع التكراري لفئات الدخل الشهري لعينة من الاسر (مئات الدولارات) في احدى المدن:

فئات الدخل الشهري (دولار)	عدد الاسر
62 – 66	3
67 – 71	8
72 – 76	20
77 – 81	21
81– 85	14
86 – 90	10
91 – 95	4

المطلوب:

1. حساب الوسط الحسابي والانحراف المعياري
2. حساب الوسيط والمنوال
3. حساب التواء التوزيع التكراري

الحـــــــل:

نكمل حساب البيانات المطلوبة باستحداث اعمدة ضمن الجدول التكراري وكما يلي:

فئات الدخل الشهري (دولار)	عدد الاسر (fi)	مركز الفئة (X_i)	(f_i x X_i)	(f_i x X_i^2)	تكرار تجمعي صاعد
62 – 66	3	64	192	**12288**	3
67 – 71	8	68	544	**36992**	11
72 – 76	20	72	1440	**103680**	31
77 – 81	21	76	1596	**121296**	52
81– 85	14	80	1120	**89600**	66
86 – 90	10	84	840	**70560**	76
91 – 95	4	88	352	**30976**	80
∑	80		6084	**465392**	

1. نحسب الوسط الحسابي والانحراف المعياري من العلاقات التالية:

$$\text{الوسط الحسابي} = \bar{X} = \frac{\sum f_i \, x \, X_i}{\sum f_i} = \frac{6084}{80} = 76.05$$

$$\text{الانحراف المعياري} = \sigma = \sqrt{\frac{\sum f_i \, x \, X_i^2}{\sum f_i} - \bar{X}^2} = \sqrt{\frac{465392}{80} - (76.05)^2} = \sqrt{33.8} = 5.8$$

2. نحسب الوسيط والمنوال من العلاقات التالية:

نوجد فئة الوسيط وذلك بقسمة مجموع التكرارات على (2) = 40 , إذن ترتيب قيمة الوسيط هي 40 , هذا الترتيب يقع ضمن الفئة الرابعة التي حديها هما(81−77).

$$\text{الوسيط} = \boldsymbol{Me = L1 + \left(\frac{\frac{N}{2} - (\sum f)}{fmedian}\right) x \, C = 76.5 + \left(\frac{\frac{80}{2} - (31)}{21}\right) x \, 5 = 78.64}$$

يتم تعيين الفئة المنوالية وهي الفئة التي تمتلك اكثر تكرارا , وهي الفئة الرابعة التي حديها هما(81−77).

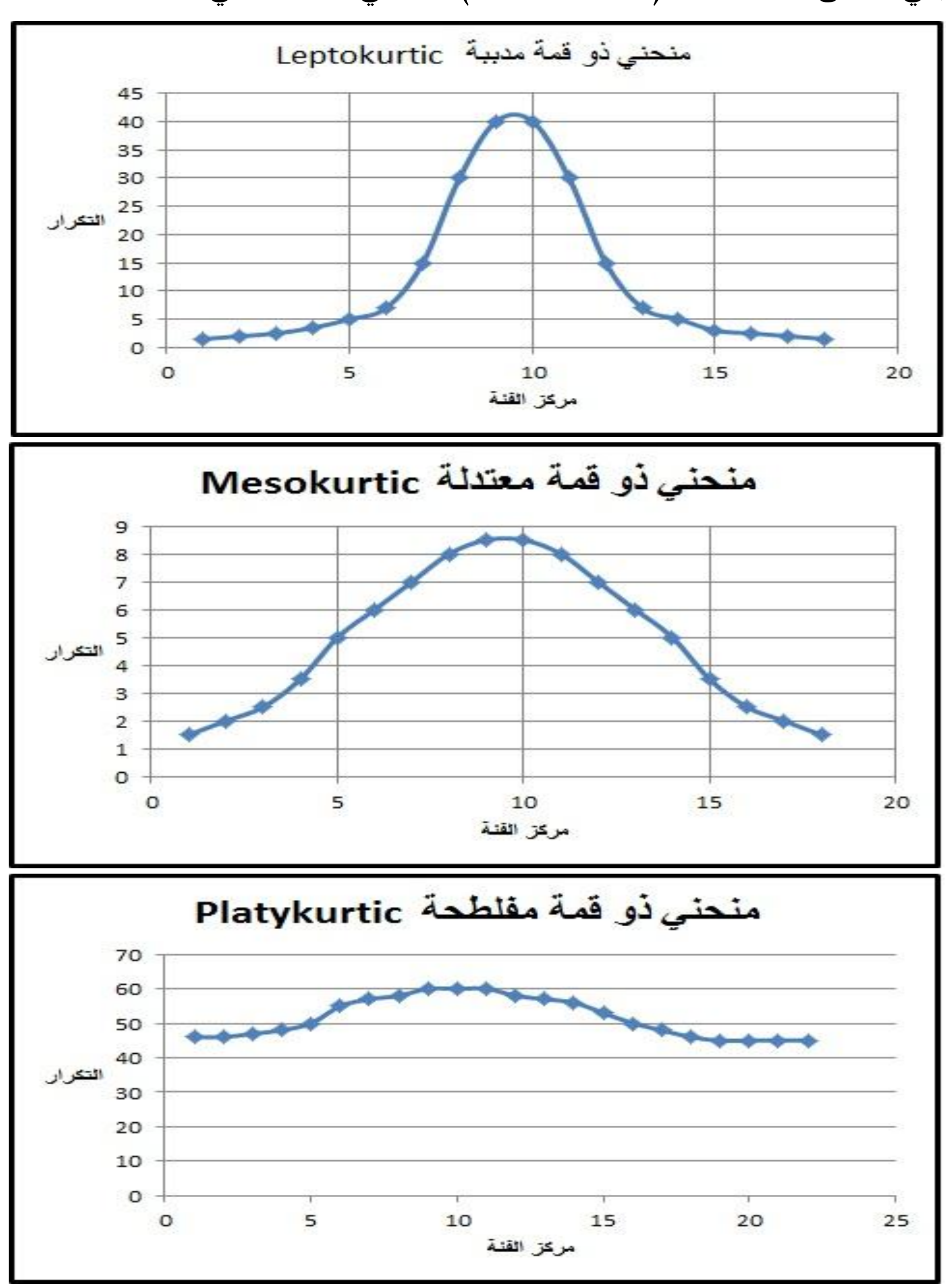
$$\text{المنوال} = \boldsymbol{Mo = L1 + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) x \, C = 76.5 + \left(\frac{1}{1+7}\right) x \, 5 = 77.12}$$

3. نحسب الالتواء باستخدام المنوال :

$$\boldsymbol{\alpha_1 = \frac{(\bar{X} - Mo)}{\sigma} = \frac{(76.05 - 77.12)}{5.8} = \frac{-1.07}{5.8} = -0.18}$$ التوزيع غير متماثل ملتوي لليسار

مقاييس التفلطح Measures of Kurtosis

التفلطح او التدبب هو انحراف قمة منحني التوزيع التكراري عن قمة المنحني الطبيعي. فالقمة العالية والضيقة حول الوسط الحسابي , تسمى قمة مدببة (Lepto–Kurtic) , أما القمة المنخفضة والمتسعة حول الوسط الحسابي, تسمى قمة مفلطحة (Platy–Kurtic) , أما القمة المعتدلة التي تشبه قمة منحني التوزيع التكراري الطبيعي فتسمى قمة معتدلة (Meso–Kurtic), كما في الشكل التالي:



اهم مقاييس التفلطح هو معامل التفلطح الذي يرمز له بالرمز $(\boldsymbol{\beta})$الذي يعرف بموجب العلاقة التالية:

$$\text{معامل التفلطح} = \frac{\text{العزم الرابع حول الوسط الحسابي}}{\text{مربع العزم الثاني حول الوسط الحسابي}} - 3 \quad , \quad \beta = \frac{m_4}{m_2^2} - 3$$

اذا كانت $\beta = 0$ سميت القمة معتدلة

اذا كانت $\beta > 0$ سميت القمة مدببة

اذا كانت $\beta < 0$ سميت القمة مفلطحة

هناك علاقات اخرى تستخدم في بعض الاحيان لايجاد معامل التفلطح واهمها هو:

$$Kurtosis = \frac{\sum(Xi - \bar{X})^4}{n\sigma^4}$$

مثال: احسب معامل التفلطح للبيانات في الجدول التكراري التالي:

الفئات Classes	التكرار fi
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
المجموع	∑ 100

الحـــل:

$$\beta = \frac{m_4}{m_2^2} - 3$$

$$m_2 = 8.5275$$

$$m_4 = 199.3759$$

$$\beta = \frac{m_4}{m_2^2} - 3 = \frac{199.3759}{8.5275} - 3 = -0.26$$

هذا يعني ان المنحني التكراري يمتلك قمة مفلطحة (Platykurtic) لان $\beta < 0$

مقاييس التشتت أو الاختلاف

Measures of Dispersion or Variation

يقصد بالتشتت أو الاختلاف بانه مقدار التباعد او الانتشار في القيم على مدى واسع ,ومقاييس التشتت يقصد بها بانها تلك القوانين او العلاقات الاحصائية التي تقيس او تحسب مدى تقارب او تباعد القيم عن وسطها الحسابي . كلما كان مقياس التشتت كبير فان ذلك يدل على وجود عدم تجانس بين القيم المستحصلة من دراسة ظاهرة معينة, وعندما يكون مقياس التشتت قليل فان ذلك يدل على ان الاختلافات بين القيم قليلة مع وجود تقارب او تجانس في القيم او القراءات. مقاييس التشتت عكس مقاييس التمركز او مقاييس النزعة المركزية حيث ان مقاييس النزعة المركزية تعطي فكرة عن مدى تقارب القرارات حول الوسط الحسابي , في حين ان مقاييس التشتت تعطي فكرة عن مدى تباعد وانتشار القيم بعيدا عن الوسط الحسابي.

ان الاعتماد على الوسط الحسابي لا يكفي عند وصف ظاهرة معينة فقد يتساوى الوسط الحسابي لمجموعتين من القراءات احداها متجانسة والاخرى غير متجانسة , لذلك لابد من اللجوء الى استخدام مقاييس التمركز ومقاييس التشتت في وصف اي ظاهرة يتم دراستها.

اهم مقاييس التشتت هي:

1. مقاييس التشتت المطلق : وهي التي تكون وحداتها نفس وحدات القيم الاصلية وهي:
   a. المدى The Range
   b. الانحراف المتوسط The Mean Deviation
   c. التباين والانحراف المعياري (القياسي) The Variance and the Standard Deviation
2. مقاييس التشتت النسبي : وهي المقاييس التي تكون خالية من وحدات القياس واهمها:
   a. معامل الاختلاف Coefficient of Variation

المدى The Range:

يعرف المدى لمجموعة من القيم او القراءات التي تنتج من دراسة ظاهرة معينة هي الفرق بين اعلى قيمة واقل قيمة في تلك المجموعة, ويرمز له بالرمز (R). في كثير من الاحيان يكون المدى مضللا ولا يعطي فكرة حقيقية عن طبيعة تجانس او توزيع القيم في المجموعة لانه يعتمد فقط على القراءتين الطرفيتين التي قد تكون شاذة في معظم الاحيان, او قد نحصل على نفس قيمة المدى لمجموعتين من القراءات احداهما ذات قيم متجانسة

والاخرى ذات قيم غير متجانسة او مشتتة. من الصعب حساب المدى في حالة الجداول التكرارية وذلك لتعذر معرفة القيمتين الطرفيتين او قيم نهايات الفئات في الجدول التكراري.

<u>مثال</u> :

اوجد المدى للمجموعة التالية من القيم: $(2, 3, 3, 5, 5, 5, 8, 10, 12)$

$$R = \text{Max. Value} - \text{Min. Value} = 12 - 2 = 10$$

<u>الانحراف المتوسط The Mean Deviation:</u>

يعرف الانحراف المتوسط بانه عبارة عن متوسط الانحرافات المطلقة (بعد اهمال الاشارة) عن الوسط الحسابي, ويرمز له بالرمز (M.D.) . الصيغة الرياضية الخاصة بحساب الانحراف المتوسط في حالة البيانات غير المبوبة او الاعداد الصحيحة هي:

$$M.D. = \frac{\sum_{i=1}^{N}|Xi - \bar{X}|}{N} = \frac{\sum|Xi - \bar{X}|}{N}$$

ان السبب في استخدام الانحرافات المطلقة (اهمال الاشارة) هو ان بقاء الاشارات الموجبة والسالبة يؤدي الى ان تكون مجموع الانحرافات يساوي صفر, وبالتالي لايمكن الحصول على نتائج حقيقية.

اوجد الانحراف المتوسط للقيم التالية: $(2, 3, 6, 8, 11)$

$$\boldsymbol{M.D.} = \frac{\sum|\boldsymbol{Xi} - \bar{\boldsymbol{X}}|}{\boldsymbol{N}} = \frac{|2-6| + |3-6| + |6-6| + |8-6| + |11-6|}{5}$$

$$= \frac{|-4| + |-3| + |0| + |2| + |5|}{5} = \frac{4+3+0+2+5}{5} = 2.8$$

في حالة الجداول التكرارية (البيانات المبوبة) فان الصيغة الرياضية لاحتساب الانحراف المتوسط سوف تكون كما يلي:

$$M.D. = \frac{\sum fi|Xi - \bar{X}|}{\sum fi}$$

حيث ان $(\boldsymbol{fi})$ تمثل التكرارات وان $(\boldsymbol{Xi})$ تمثل مراكز الفئات

مثال:

أوجد الانحراف المتوسط من الجدول التكراري التالي:

Classes	fi	Xi	fi *x* Xi	$\lvert Xi - \bar{X}\rvert$	fi $\lvert Xi - \bar{X}\rvert$
60 – 62	5	61	305	6.45	32.25
63 – 65	18	64	1152	3.45	62.10
66 – 68	42	67	2814	0.45	18.90
69 – 71	27	70	1890	2.55	68.85
72 – 74	8	73	584	5.55	44.40
المجموع	$\sum 100$		$\sum 6745$		$\sum 226 > 5$

ملاحظة: تهمل الاشارة عن وضع البيانات بين خطين مستقيمينيدل على ان البيانات مطلقة.

$$\bar{X} = \frac{\sum fi\ x\ Xi}{\sum fi} = \frac{6745}{100} = 67.45$$

$$M.D. = \frac{\sum fi|Xi-\bar{X}|}{\sum fi} = \frac{226.5}{100} = 2.265$$

التباين والانحراف القياسي Variance and Standard Deviation :

لاحظنا في الانحراف المتوسط انه يمكن التغلب على مشكلة الاشارة السالبة التي تؤدي الى ان يكون مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر باستخدام القيم المطلقة للانحرافات, اي استخدام الارقام بدون اشارات. يمكن التغلب غلى هذه المشكلة بطريقة اخرى اسهل واسرع وذلك بتربيع قيم الانحرافات التي تؤدي الى ان تصبح جميع القيم موجبة, وبذلك نحصل على مجموع مربعات الانحرافات (Sum of Squares) والتي يرمز لها (.S.S) والتي تساوي :

$$S.S. = \sum (Xi - \bar{X})^2$$

في حالة قسمة مجموع مربعات القيم على عدد القيم او القراءات وبذلك نحصل على التباين الذي يرمز له بالرمز $(S)^2$ او $(\sigma)^2$, وتكون الصيغة الرياضية لحساب التباين هو:

$$\sigma^2 = \frac{\sum (Xi - \bar{X})^2}{N}$$

لكي ناخذ بنظر الاعتبار حجم العينة او الظاهرة المدروسة وحتى يمكن اجراء مقارنة بين العينات المختلفة الاحجام يتم قسمة مجموع مربع الانحرافات على (n−1), حيث يتم سحب او استبعاد احد النماذج من المجموعة حتى لا يكون مجموع الانحرافات عن السط الحسابي يساوي صفر, وهو ما يطلق عليه بدرجة الحرية,

وبذلك يكون عدد القيم الحرة في اية عينة هي (n−1) وهو ما يسمى درجة الحرية. تستخدم درجة الحرية في العينة او الظاهرة المدروسة عندما يكون لدينا عدد العينات او القيم اكثر من (30) قيمة حيث ان سحب او استبعاد احد القيم لا يؤثر عل النتيجة النهائية لحساب التباين. وبذلك تكون الصيغة الرياضية لحساب التباين هي كالاتي:−

عند حساب التباين تم استخدام مربعات الانحرافات, وبذلك فان قيمة التباين بهذه الحالة تكون مقاسة بمربع الوحدات المستخدمة في قياس القيم او القراءات, مثلا اذا كانت القيم مقاسة بالمتر فان التباين يكون مقاسا المتر المربع, وبهذه الحالة لا توجد مشكلة في استخدام وحدات القياس, ولكن المشكلة تظهر عندما تكون وحدات القياس المستخدمة هي عبارة عن الوزن بالكيلوغرام مثلا أو مبالغ مالية بالدينار او عدد اطفال الاسر المختلفة, فان التباين يكون مقاسا بالكيلوغرام المربع أو الدينار المربع او الطفل المربع وهذه جميعا غير ذات معنى أو غير معقولة, لغرض حل مثل هذا الاشكال يجب اعادة وحدات القياس الى اصلها والطريقة هي استخدام الجذر التربيعي للتباين $(\sigma)^2$ لنحصل على (σ) وهو ما يطلق عليه بالانحراف القياسي او الانحراف المعياري (Standard Deviation), والذي يكون مقاسا بالوحدات الاصلية أو نفس الوحدات التي تستخدم في قياس القيم او القراءات الاصلية. يعرف الانحراف المعياري بانه الجذر التربيعي لمتوسط مربعات انحرافات (الفرق بين القيم عن وسطها الحسابي ويرمز له بالرمز (S) في حالة البيانات غير المبوبة والرمز (σ) في حالة البيانات المبوبة. الصيغة الرياضية المستخدمة في ايجاد الانحراف القياسي هي:−

$$\text{الانحراف القياسي} = S = \sqrt{\frac{\sum(Xi-\bar{X})^2}{n-1}} = \sqrt{\frac{\sum(Xi)^2}{n-1} - (\overline{Xi})^2}$$

هنالك طريقتين لايجاد الانحراف القياسي في حالة البيانات غير المبوبة او الاعداد الصحيحة هما:

A. الطريقة المطولة: تستخدم فيها الصيغة الرياضية التالية:

$$\text{الانحراف القياسي} = S = \sqrt{\frac{\sum(Xi-\bar{X})^2}{n-1}} = \sqrt{\frac{\sum(Xi)^2}{n-1} - (\overline{Xi})^2}$$

B. الطريقة المختصرة تستخدم فيها الصيغة الرياضية التالية:

$$\text{الانحراف القياسي} = S = \sqrt{\frac{\sum xi^2 - \frac{(\sum Xi)^2}{n}}{n-1}}$$

الفرق بين الطريقة المطولة والطريقة المختصرة هي في سهولة وسرعة ايجاد المتغيرات المطلوبة لغرض تعويضها في المعادلة الرياضية.

مثال(1): أوجد الانحراف القياسي للاعداد او القيم التالية: (5 ,18 ,10 ,15 ,3 ,7 ,6 ,12) باستخدام الطريقة المطولة والطريقة المختصرة.

الحل: نستخرج الوسط الحسابي للقيم اعلاه:

$$\bar{X} = \frac{\sum Xi}{N} = \frac{12+6+7+3+15+10+18+5}{8} = \frac{76}{8} = 9.5$$

A. نستخرج الانحراف القياسي باستخدام الطريقة المطولة:

$$S = \sqrt{\frac{\sum(Xi - \bar{X})^2}{n-1}}$$

$$= \sqrt{\frac{(12-9.5)^2+(6-9.5)^2+(7-9.5)^2+(3-9.5)^2+(15-9.5)^2+(10-9.5)^2+(18-9.5)^2+(5-9.5)^2}{8}}$$

$$= \sqrt{\frac{6.25+12.25+6.25+42.25+30.25+0.25+72.25+20.25}{8}} = \sqrt{\frac{190}{8}} = \sqrt{23.75} = 4.8$$

B. نستخرج الانحراف القياسي باستخدام الطريقة المختصرة:

$$S = \sqrt{\frac{\sum xi^2 - \frac{(\sum Xi)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{([144+36+49+9+225+100+324+25]) - \frac{(12+6+7+3+15+10+18+5)^2}{8}}{8}}$$

$$= \sqrt{\frac{912 - \frac{5776}{8}}{8}} = \sqrt{\frac{190}{8}} = \sqrt{23.75} = 4.8$$

ملاحظة: لم يتم طرح احدى القراءات من مجموع هذه القراءات كون عدد القراءات اقل من 30 قراءة.

مثال (2): اوجد الانحراف المعياري للاجور اليومية بالدولار للعينة التالية المكونة من خمس عمال باحدى الشركات : (50, 70, 80, 90, 60).

الحل: 1- نحسب الوسط الحسابي : $\bar{X} = \frac{\sum Xi}{N} = \frac{60+90+80+70+50}{5} = 70\$$

2- نحسب الانحراف المعياري من احدى العلاقات المعطاة سابقا ولتكن العلاقة التالية:

$$\sqrt{\frac{\sum(Xi)^2}{n-1} - (\bar{X})^2} = \frac{\sqrt{(60)^2+(90)^2+(80)^2+(70)^2+(50)^2}}{5} - (70)^2 = \sqrt{5100-4900} = 14.1\$$$

في حالة البيانات المبوبة أو الجداول التكرارية فيمكن ايجاد الانحراف القياسي ايضاً بطريقتين هما :

A. الطريقة المطولة ويستخدم فيها القانون التالي:

$$\sigma = \sqrt{\frac{\sum fi(Xi - \bar{X})^2}{\sum fi - 1}}$$

B. الطريقة المختصرة ويستخدم فيها القانون التالي:

$$\sigma = \sqrt{\frac{\sum(fi\,Xi^2) - \frac{(\sum fi\,Xi)^2}{\sum fi}}{\sum fi - 1}}$$

مثال(1) : احسب الانحراف القياسي بالطريقة المطولة والطريقة المختصرة من جدول التوزيع التكراري التالي:

الفئات	التكرار
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 - 74	8
المجموع	100

الحـــــل:

A. الطريقة المطولة : نستخرج المتغيرات التي نحتاجها من الجدول التكراري لغرض تعويضها في القانون:

الفئات	التكرار fi	مركز الفئة Xi	fi x Xi	$(Xi-\bar{X})$	$(Xi-\bar{X})^2$	$fi(Xi-\bar{X})^2$
60 – 62	5	61	305	-6.45	41.6025	208.0125
63 – 65	18	64	1152	-3.45	11.9025	214.2450
66 – 68	42	67	2814	-0.45	0.2025	8.5050
69 – 71	27	70	1890	2.55	6.5025	175.5675
72 - 74	8	73	584	5.55	30.8025	246.4200
المجموع	100		6745			852.75

نستخرج قيمة الوسط الحسابي:

$$\bar{X} = \frac{\sum fi\,x\,Xi}{\sum fi} = \frac{6745}{100} = 67.45$$

$$\sigma = \sqrt{\frac{\sum(Xi-\bar{X})^2}{n-1}} = \sqrt{\frac{852.75}{100-1}} = \sqrt{8.6} = 2.9$$

B. الطريقة المختصرة نستخرج المتغيرات المطلوبة ثم نطبق القانون الخاص بها:

الفئات	التكرار fi	مركز الفئة Xi	fi x Xi	$(Xi)^2$	$(fi \times Xi^2)$
60 – 62	5	61	305	3721	18605
63 – 65	18	64	1152	4096	73728
66 – 68	42	67	2814	4489	188538
69 – 71	27	70	1890	4900	132300
72 - 74	8	73	584	5329	42632
المجموع	100		6745		455803

$$\sigma = \sqrt{\frac{\sum(fi\,Xi^2) - \frac{(\sum fi\,Xi)^2}{\sum fi}}{\sum fi - 1}} = \sqrt{\frac{455803 - \frac{(6745)^2}{100}}{99}}$$

$$= \sqrt{\frac{455803 - 454950.25}{99}} =$$

$$\sqrt{\frac{852.75}{99}} = \sqrt{8.6} = 2.9$$

مثال (2): أوجد الانحراف المعياري لاجور العمال حسب الجدول التكراري التالي:

الفئات (اجور العمال)	التكرار (عدد العمال)(f_i)
10 – 19	3
20 – 29	6
30 – 39	10
40 – 49	15
50 – 59	8
60 – 69	5
70 – 79	3

الحـــل:

1. يتم عمل جدول تكراري نستحث فيه اعمدة جديدة لغرض استخراج وحساب المتغيرات المطلوبة.

الفئات (اجور العمال)	التكرار (f_i)(عدد العمال)	مركز الفئة (X_i)	$(f_i \times X_i)$	$(f_i \times X_i^2)$
10 – 19	3	14.5	43.5	630.75
20 – 29	6	24.5	147	3601.5
30 – 39	10	34.5	345	11902.5
40 – 49	15	44.5	667.5	29703.75
50 – 59	8	54.5	436	23762
60 – 69	5	64.5	322.5	20801.25
70 – 79	3	74.5	223.5	16650.75
$\sum$	50		2185	107052.5

2. يتم حساب الوسط الحسابي من العلاقة التالية:

$$\bar{X} = \frac{\sum f_i \; x \; X_i}{\sum f_i - 1} = \frac{2185}{49} = 44.59$$

3. نستخدم احدى العلاقات في استخدام الانحراف المعياري اما الطريقة المطولة او الطريقة المختصرة :

$$\sigma = \sqrt{\frac{\sum(f_i \; x \; X_i^2)}{\sum f_i - 1} - \bar{X}^2} = \sqrt{\frac{107052.5}{49} - (44.59)^2} = \sqrt{2184.7 - 1988.2} = \sqrt{196.5} = 14$$

خواص التباين والانحراف القياسي

1– ان متوسط مربعات انحرافات (فروق) القيم عن وسطها الحسابي يسمى التباين, أي إن الانحراف المعياري هو جذر التباين.

2– قيمة التباين لا بد أن تكون موجبة أو تساوي صفر.

3– كلما اقتربت قيمة التباين من الصفر , أي كلما اقتربت قيمة الانحراف المعياري من الصفر , كلما اصبحت البيانات او القراءات في الظاهرة المدروسة قريبة من التجانس.

4– يتاثر الانحراف المعياري بالقيم الشاذة.

5– عند اضافة أو طرح اي عدد ثابت (نرمز له بالرمز K) الى كل قيمة من القيم او القراءات , فان قيمة التباين والانحراف القياسي تبقى كما هي لا تتغير . اي ان :

قيمة التباين الجديدة = قيمة التباين الاصلية

قيمة الانحراف القياسي الجديدة = قيمة الانحراف القياسي الاصلية

<u>مثال</u> : احسب التباين والانحراف القياسي للقيم التالية اولاً , ثم اضف لكل منهما (3) التباين والانحراف القياسي الجديدة. (8, 3, 2, 12, 10) .

a) نحسب قيمة التباين والانحراف القياسي قبل الاضافة:

$$\text{الانحراف القياسي} = \boldsymbol{\sigma} = \sqrt{\frac{\sum \boldsymbol{xi^2} - \frac{(\sum \boldsymbol{Xi})^2}{\boldsymbol{n}}}{\boldsymbol{n-1}}}$$

$$= \sqrt{\frac{(\mathbf{8})^2 + (\mathbf{3})^2 + (\mathbf{2})^2 + (\mathbf{12})^2 + (\mathbf{10})^2 - \frac{(\mathbf{8+3+2+12+10})^2}{\mathbf{5}}}{\mathbf{4}}}$$

$$= \sqrt{\frac{\mathbf{76}}{\mathbf{4}}} = \sqrt{\mathbf{19}} = \mathbf{4.3}$$

∴ التباين = 19 , الانحراف القياسي = 4.3

b) نحسب قيمة التباين والانحراف القياسي بعد الاضافة:

بعد اضافة ثابت قيمته (3) تصبح القيم كما يلي: (13 ,15 ,5 ,6 ,11), ثم نحسب :

$$\text{الانحراف القياسي} = \sigma = \sqrt{\frac{\sum xi^2 - \frac{(\sum Xi)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{(11)^2 + (6)^2 + (5)^2 + (15)^2 + (13)^2 - \frac{(11+6+5+15+13)^2}{5}}{4}} = \sqrt{\frac{76}{4}}$$

$$= \sqrt{19} = 4.3$$

نحصل على نفس النتيجة في حالة اضافة او طرح عدد ثابت من القيم او القراءات الاصلية.

6- اذا ضربت كل قيمة من القيم الاصلية بعدد ثابت (K) فان :

التباين للقيم الجديدة = التباين للقيم الاصلية x مربع العدد الثابت

الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية x العدد الثابت

<u>العلاقة بين الانحراف القياسي والانحراف المتوسط</u>

A. اذا كان التوزيع التكراري غير متماثل (المنحني التكراري ملتوي التواء بسيط) فان:

الانحراف المتوسط = $\frac{4}{5}$ الانحراف القياسي, $\quad \boldsymbol{M.D.} = \frac{\mathbf{4}}{\mathbf{5}} \boldsymbol{\sigma}$

B. عند قياس مدى تشتت متوسط العينات اتابعة لظاهرة معينة فانه يسخدم ما يسمى بالخطا القياسي (Standard Error) او الانحراف القياسي للمتوسطات (Standard Deviation of the Mean) ويرمز له بالرمز $(S_{\overline{x}})$, ويحسب بالقانون التالي:

$$(S_{\overline{x}}) = \sqrt{\frac{\sigma^2}{n}}$$

<u>مقاييس التشتت النسبي :</u>

هي المقاييس التي تكون خالية من وحدات القياس, وهي ذات اهمية كبيرة عند اجراء مقارنة بين تشتت مجموعتين او اكثر تختلف في وحدات القياس لقيمها. ومن اهم مقاييس التشتت النسبي هي:

1. <u>معامل الاختلاف Coefficient of Variation:</u>

وهو معامل اختلاف نسبي يستخدم للمقارنة بين تشتت ظاهرتين او اكثر مختلفتين او حتي متشابهتين في وحدة القياس. ان الظاهرة التي يكون او تمتلك معامل اختلاف اكبر تكون اكثر تشتتاً من الاخرى.

c) <u>يمكن حساب معامل الاختلاف بدلالة الانحراف القياسي والوسط الحسابي:</u>

$$C.V. = \frac{\sigma}{\overline{\overline{X}}} \, x\, 100$$

d) <u>يمكن حساب معامل الاختلاف بدلالة الانحراف المتوسط والوسط الحسابي:</u>

$$C.V. = \frac{M.D.}{\overline{X}} \, x\, 100$$

e) <u>يمكن حساب معامل الاختلاف بدلالة الانحراف المتوسط والوسيط:</u>

$$C.V. = \frac{M.D.}{Me} \, x\, 100$$

مثال (1):

أوجد معامل الاختلاف من الجدول التكراري التالي:

الفئات (اجور العمال)	التكرار (عدد العمال)(f_i)
10 – 19	3
20 – 29	6
30 – 39	10
40 – 49	15
50 – 59	8
60 – 69	5
70 – 79	3

الحـــل:

A. نستخرج الوسط الحسابي والانحراف المعياري من الجدول التكراي وكما يلي:

$$\bar{X} = 44.59 \quad , \qquad \sigma = 14$$

B. نستخرج معامل الاختلاف بدلالة الوسط الحسابي والانحراف المعياري وحسب العلاقة التالية:

$$\boldsymbol{C.V. = \frac{\sigma}{\bar{X}}\ x\ 100} = \frac{14}{44.59} x\ 100 = 33.39\%$$

مثال (2): اذا كانت نتائج الامتحانات النهائية في مادتي الاحصاء والصخور للمرحلة الثانية كما يلي:

المتغير / المادة	الاحصاء	الصخور
الوسط الحسابي	78	73
الانحراف القياسي	80	76

للاحصاء $$\boldsymbol{C.V. = \frac{\sigma}{\bar{\bar{X}}}\ x\ 100 = \frac{80}{78}\ x\ 100 = 102.5\%}$$

للكيمياء $$\boldsymbol{C.V. = \frac{\sigma}{\bar{\bar{X}}}\ x\ 100 = \frac{76}{73}\ x\ 100 = 104.1\%}$$

من ذلك نستنتج ان درجة التشتت في مادة الكيمياء اكثر عما هي في الاحصاء

2. الدرجة القياسية Standardized Scores

في معظم الاحيان وخلال مراحل المعالجات الاحصائية عند دراسة عدة ظواهر معينة, فاننا نحتاج الى مقارنة ظاهرتين او مجموعتين مختلفتين , لكل ظاهرة من هذه الظواهر لها وحدات القياس الخاصة بها, لذلك يجب تحويل وحدات كل ظاهرة او مجموعة الى وحدات قياسية لكي نتمكن من اجراء عملية المقارنة وتكوم ذات

معنى . يتم ذلك باستخدام الوسط الحسابي والانحراف القياسي لكل ظاهرة او مجموعة. يتم ذلك باستخدام العلاقة التاليـــــة:

$$\text{الدرجة القياسية} = Zi = \frac{Xi - \bar{X}}{\sigma}$$

مثـــال:

حصل احد الطلاب على درجة (84) في الامتحان النهائي في مادة الرياضيات, وكان الوسط الحسابي لجميع طلبة المرحلة في الرياضيات تساوي (76) وبانحراف قياسي قدره (10). أما في امتحان الفيزياء فقد حصل نفس الطالب على درجة (90) وكان الوسط الحسابي لجميع طلبة المرحلة هو (82) وبانحراف قياسي قدره (16). في اي الموضوعين كانت قابلية الطالب اعلى؟

الحــــل:

نحول درجات المواد في الفيزياء والرياضيات الى درجات قياسية لغرض التخلص من وحدات القياسباستخدام قانون الدرجة القياسية:

بالنسبة للرياضيات: $Zi = \frac{Xi - \bar{X}}{\sigma} = \frac{84-76}{10} = 0.8$

بالنسبة للفيزياء : $Zi = \frac{Xi - \bar{X}}{\sigma} = \frac{90-82}{16} = 0.85$

من هذا يتضح بان قابلية الطالب في الرياضيات اعلى عما هي عليه في الفيزياء بالرغم من ان درجته في الرياضيات اقل عما هي في الفيزياء.

مســـــائل اثرائيــــــــة:

مسالة (1):

الجدول التكراري التالي يوضح رواتب عينة من (100) موظف بالدينار .

فئات الرواتب (دينار 1000X)	التكرار (عدد الموظفين)
30 – 39	4
40 – 49	11
50 – 59	20
60 – 69	36
70 – 79	17
80 – 89	8
90 - 99	4
∑	100

المطلوب:

A. ارسم المدرج التكراري لتوزيع الرواتب , ثم حدد منه قيمة المنوال.

B. ارسم المنحني المتجمع الصاعد والنازل , ثم اوجد منه قيمة الوسيط في الرواتب.

C. من المنحني التكراري التجمعي الصاعد , اوجد عدد الموظفين الذين تقل رواتبهم عن 750 الف دينار

مسالة (2):

الجدول التكراري التالي يوضح توزيع عينة مكونة من (200) موظف باحدى الشركات موزعين حسب اعمارهم بالسنة:

فئات العمر (سنة)	التكرار (عدد الموظفين)
20 – 24	10
25 – 29	17
30 – 34	24
35 – 39	43
40 – 44	34
45 – 49	30
50 – 54	23
55 – 59	19
∑	200

المطلوب:

1. احسب الوسط الحسابي والانحراف المعياري.
2. ارسم المنحني التجمعي الصاعد ومن الرسم اوجد :
a) نسبة عدد المشتغلين الذين تقل اعمارهم عن 40 سنة.
b) الحد الاعلى للعمر الذي بلغه 120 موظف.
3. الوسيط باستخدام التمثيل الصوري.
4. ارسم المنحني التكراري, هل المنحني متماثل؟ دلل على الاجابة بحساب معامل الالتواء.

مسالة (3):

الجدول التالي يوضح توزيع عينة من الموظفين في احدى الشركات الخاصة موزعين حسب فئات الراتب:

عدد الموظفين	فئات الراتب
3	60 – 64
8	65 – 69
20	70 – 74
25	75 – 79
10	80 – 84
10	85 – 89
4	90 - 94
80	$\sum$

المطلوب:

1. احسب الوسيط باستخدام طريقيتين (الحساب الرياضي والتمثيل الصوري بارسم).
2. المنوال باستخدام طريقتين (الحساب الرياضي والتمثيل الصوري بارسم).
3. معامل الاختلاف.
4. معامل الالتواء الاول والثاني.

مسالة (4):

البيانات التالية تمثل اسعار ستة من المنتجات الغذائية (بالدينار) مختارة من ثلاث مراكز تجارية وكما يلي:

عينة (1) : (7, 2, 1, 3, 6, 5)

عينة (2) : (8, 4, 2, 1, 3, 6)

عينة (3) : (1, 4, 2, 5, 1, 3)

احسب معامل الاختلاف لكل عينة , أي العينات اكثر تشتتاً.

مسالة (5):

الجدول التالي يوضح توزيع عينة من (500) موظف موزعين حسب ساعات العمل التي يحققونها يومياً في احدى شركات القطاع الخاص:

ساعات العمل	عدد الموظفين
2 – 3	98
4 – 5	118
6 – 7	101
8 – 9	95
10 – 11	50
12 – 13	20
14 – 15	10
16 - 17	8
$\sum$	500

المطلوب:

1. احسب الوسيط والمنوال
2. احسب الانحراف المعياري
3. احسب معامل الاختلاف
4. ارسم المنحني التكراري, هل ان التوزيع متماثل؟ علل الاجابة؟

مسالة (6):

من الجدول التكراري التالي :

الفئات	التكرار
15 – 17	3
18 – 20	7
21 – 23	21
24 – 26	5
27 – 29	4

اوجد:

1. الانحراف المعياري
2. الوسيط

الفصل السابع

القطاعات الدائرية (المساحات الدائرية) Pie Charts :

وهي عبارة عن دائرة تقسم الى قطاعات زواياها المركزية تتناسب مع القراءات , أو بصيغة اخرى يتم تقسيم الدائرة الى قطاعات بنسبة القيم الظاهرة وبحسب القيم الظاهرة وبحسب قياس زاوية كل قطاع نسبة الى الزاوية المركزية ($360°$) , ويمكن حساب الزاوية الخاصة بكل قطاع القراءة الخاصة به وكالاتي :

$$\text{زاوية القطاع} = \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^{\circ}$$

او بصيغة اخرى :

$$\text{الزاوية المركزية} = \frac{\text{القراءة نفسها}}{\text{مجموع القراءات}} \; x \; \text{الزاوية المركزية}$$

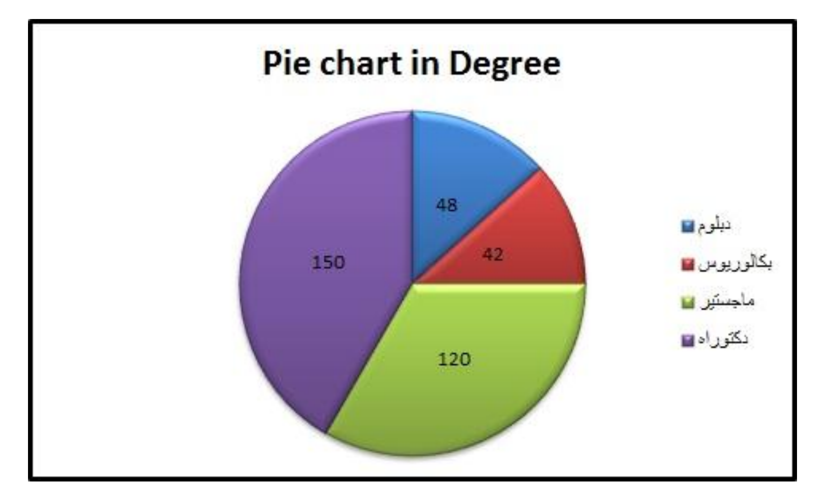
مثال (1): البيانات المدجة في الجدول التالي تمثل مؤهلات اعضاء هيئة التدريس في كلية العلوم , مثل هذه البيانات بالقطاعات الدائرية .

المؤهل	دبلوم	بكالوريوس	ماجستير	دكتوراه	
العدد	8	7	20	25	$\sum 60$

الحل:

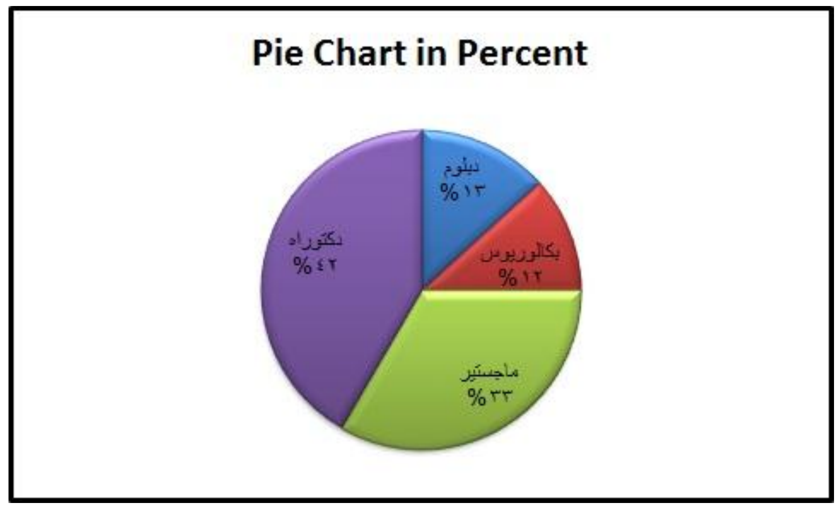
نرسم الدائرة ثم نحدد الزاوية المركزية لكل قطاع من العلاقة الرياضية وكما يلي :

1. $\text{زاوية القطاع} = \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^{\circ} \; , \; \text{الدبلوم} = \frac{8}{60} \; x \; 360^{\circ} = 48$
2. $\text{زاوية القطاع} = \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^{\circ} \; , \; \text{البكالوريوس} = \frac{7}{60} \; x \; 360^{\circ} = 42$
3. $\text{زاوية القطاع} = \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^{\circ} \; , \; \text{الماجستير} = \frac{20}{60} \; x \; 360^{\circ} = 120$
4. $\text{زاوية القطاع} = \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^{\circ} \; , \; \text{الدكتوراه} = \frac{25}{60} \; x \; 360^{\circ} = 150$



نلاحظ هنا ان مجموع الزوايا لا بد ان يساوي (360)درجة = (42+33+12+13)

وممكن ان نرسم القطاعات الدائرية بالنسبة المئوية , حيث يتم تقسيم درجة كل قطاع على (°360) وتضرب في (100) , وكما يلي : $\text{نسبة كل قطاع} = \frac{\text{درجة كل قطاع}}{360} \, x \, .100$



مثال (2): الجدول التالي يمثل تقريبا مساحات القارات في العالم , مثلها بالقطاعات او الرسوم الدائرية بالدرجات وبالنسبة المئوية .

القارة	المساحة بالمليون كم^2
اسيا	47
امريكا الجنوبية	18
افريقيا	30
استراليا ونيوزلندا	8
اوربا	5
امريكا الشمالية	24
المجموع	132

الحل : (a) : نرسم الدائرة ثم نحدد الزاوية المركزية لكل قطاع من العلاقة الرياضية وكما يلي :

1. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, اسيا $= \frac{47}{132} \; x \; 360^\circ = 128$

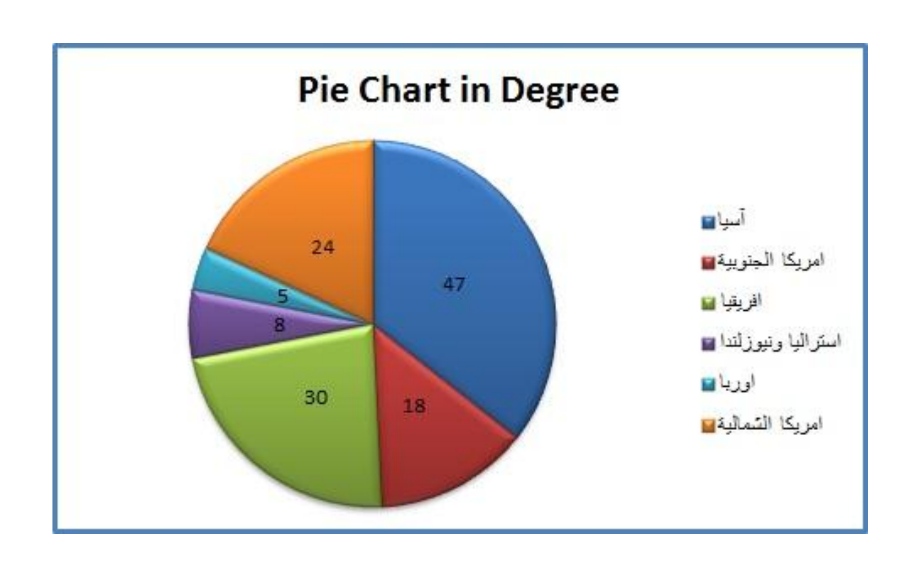
2. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, امريكا الجنوبية $= \frac{18}{132} \; x \; 360^\circ = 49$

3. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, افريقيا $= \frac{30}{132} \; x \; 360^\circ = 82$

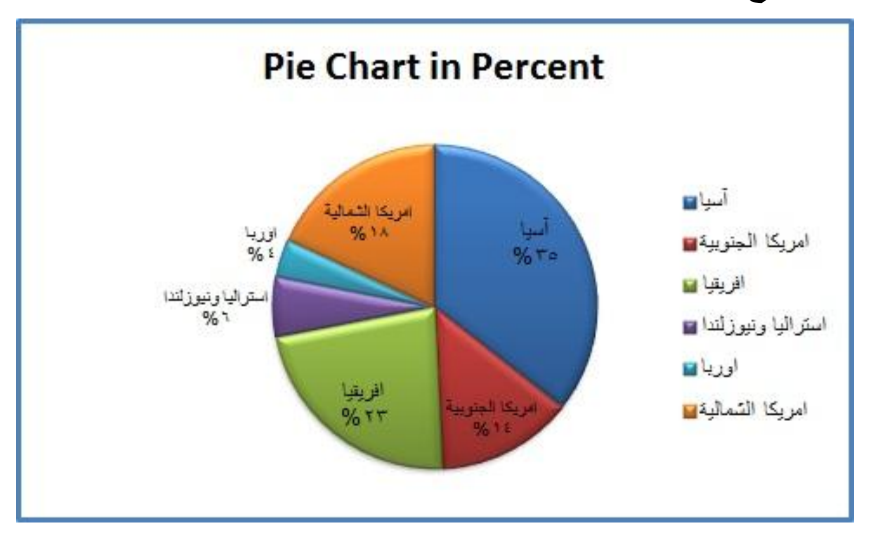
4. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, استراليا ونيوزيلندا $= \frac{8}{132} \; x \; 360^\circ = 22$

5. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, اوربا $= \frac{5}{132} \; x \; 360^\circ = 14$

6. زاوية القطاع $= \frac{\text{قيمة (تكرار) الجزء الممثل بالقطاع}}{\text{مجموع القيم (التكرارات)}} \; x \; 360^\circ$, امريكا الشمالية $= \frac{24}{132} \; x \; 360^\circ = 65$



(b) : نرسم مساحة كل قطاع حسب النسبة المئوية :



الباب الثاني

الاحصاء الاستدلالي

نظرية الاحتمال

الفصل الثامن

نظرية الاحتمال

تمهيد:

تعرف نظرية الاحتمال بانها فرع من فروع الرياضيات التطبيقية التي تهتم بدراسة التجارب العشوائية, او بمعنى آخر إنها تهتم بدراسة تاثير الصدفة على الظواهر والاشياء.

تلعب الاحتمالات دوراً كبيرا في حياتنا اليومية وتدخل في كثير من العلوم الاخرى التي تعتمد على اتخاذ القرارات بطريقة الاحتمال أو عند عدم التاكد من النتيجة, مثال على ذلك قد نلغي رحلة خارجية تم الترتيب لها منذ فترة طويلة وذلك لاحتمال ان يكون الطقس ردئ أو لاحتمال سقوط المطر, أو يمكن للطالب ان يهمل قراءة جزء من المنهج الدراسي لاحتمال ان لا ياتي في الامتحان. هذه التقديرات تستند على اساس رياضي ولكنها تعتمد على الخبرة المتراكمة , مثلا احوال الطقس أو الاسئلة الامتحانية المرشحة في الامتحانات.

نشأت نظرية الاحتمال في القرن السابع عشر من خلال الابحاث التي قام بها كل من العالم باسكال (Pascal 1662 – 1623) والعالم فيرمات (Fermat 1665 – 1601) عند دراستهم لارقام معينة في عالم المراهنات , ثم اكمل برنولي (Bernoulli 1705 – 1645) وبعده جاء لابلاس (Laplace 1827 – 1749) حيث وضع تعريف واسس نظرية الاحتمال.

هنالك تطبيقات عديدة طبقت فيها نظرية الاحتمال مثل الارصاد الجوية, العلوم الهندسية ونظرية الوراثة التي جاء بها العالم مندل (Mendel). أولى التطبيقات لنظرية الاحتمال بنيت على التجارب التالية:

1. إذا القينا قطعة من المعدن في الهواء فانها سوف تسقط على الارض وهذا شئ (مؤكد) لانها حقيقة معروفة , ولكن اذا القينا قطعة نقود على طاولة مسطحة , فان قطعة النقود سوف تسقط وتستقر على احد وجهيها (لا يمكن للقطعة النقدية ان تستقر على حافتها) , ولكننا قبل اجراء رمي القطعة النقدية لا نعلم اي الوجهين سوف يظهر الى الاعلى, وهذا يعتمد على ما نطلق عليه (بالصدفة) أو الاحتمال.
2. مثال آخر , اننا نعرف ان الماء يتحول الى بخار اذا تم تسخينه الى درجة حرارة (100 C°) وهذا شئ مؤكد , ولكن عند رمي زهرة الطاولة (زهرة النرد) على لوح فان ما نعرفه هو ان احد الاوجه الستة سوف يظهر الى الاعلى , ولكننا لا نعرف اي من الاوجه السته سوف يظهر الى الاعلى, لان ذلك يعتمد على (الصدفة).

من خلال هذه الامثلة يمكن ان نفرق بين الشئ (المؤكد), و(الصدفة) , فالاول يدل على شيء معلوم لدينا وان نتيجة التجربة تحدث وان كل الظروف تكون مؤاتية تؤدي الى حدوثه. أما الثاني فاننا لا نعرف أو لا يمكن التكهن بالنتيجة وانها تدل على شيء غير معلوم لدينا تماما , ولا نعرف الظروف التي تؤدي الى حدوث النتيجة الا بعد اجراء التجربة. مثال آخر على مثل هذه الحالة عندما نقول ان (السماء يحتمل ان تمطر اليوم) يمكن ان نطلق كلمة احتمال على ان السماء يحتمل ان تمطر اعتمادا على توفر بعض الظروف التي تؤدي الى سقوط المطر. مثلا اذا كانت السماء ملبدة بالغيوم والجو مائلا الى البرودة ولكن هي ليست كل الظروف حيث توجد ظروف اخرى لا نعرفها تماماً اذا ما توفرت جميعا فان المطر سوف يسقط بصورة مؤكدة وعند عدم توفرها جميعا هناك احتمال بعدم سقوط المطر.

كذلك يمكن النظر الى الاحتمالات على انها احد فروع الرياضيات الذي سبق ان اشرنا اليه في بداية تعريف نظرية الاحتمال وهي النظرية التي تهتم بدراسة التجارب او المحاولاتالعشوائية, وتسمى التجربة أو المحاولة (عشوائية) اذا كانت من غير الممكن تحديد النتيجة بالضبط , أي لا نستطيع التنبؤ بها الا بعد اجراء التجربة.

مثلاَ: اذا تم القاء قطعة معدنية أو قطعة نقود فاننا لا نستطيع ان نتنبأ بالنتيجة اذا كان السطح العلوي لها سيظهر لنا صورة (Head) أو كتابة (Tail) , اذاَ هذه هي تجربة عشوائية لاننا لم نتمكن من معرفة النتيجة بالضبط الا بعد اجراء التجربة. كذلك عند سحب ورقة اللعب عشوائياً من كجكوعو اوراق اللعب بعد خلطها جيداً, فاننا لا نعلم اذا كانت الورقة المسحوبة ستظهر صورة او عدد معين, كذلك اذا كانت هنالك حالة ولادة فلا نستطيع التنبؤ اذا كان المولود ذكراً ام انثى, اذا هذه هي الاسس والتعاريف التي تقوم عليها التجربة العشوائية.

<u>شروط التجربة العشوائية:</u>

1. يمكن للتجربة العشوائية ان تصف جميع النتائج التي من الممكن وقوعها او حدوثها قبل عملية اجراء التجربة او المحاولة .
2. لا يمكن بالضبط تحديد نتيجة التجربة الا بعد اجراء التجربة أو اجراء عملية المحاولة.

عند رمي زهرة الطاولة أو حجر النرد (Dice) مرة واحدة, وملاحظة الوجه الذي سوف يظهر لنا, نحن نعرف مسبقا ان الوجه الذي يظهر من رمية واحدة ممكن ان يكون احد الارقام التالية: (1,2,3,4,5,6) , اي بهذه الحالة يمكن تحديد او تشخيص جميع الحالات الممكنة او التي من الممكن ان تظهر لنا احد هذه الارقام,

ولكننا لا نتمكن من معرفة النتيجة بالضبط او اي من الارقام سوف يظهر على الوجه السطحي او العلوي الا بعد اجراء التجربة بعينها, ولذلك سميي هذا النوع من التجارب بالتجارب العشوائية.

هنالك بعض المصطلحات والتعاريف التي وضعت لغرض وصف التجارب العشوائية وهي كما يلي:

1. فضاء العينة Sample Space :

ويرمز لها بالرمز (A) , وهي تمثل جميع النتائج الممكنة التي يمكن ان تحصل في اي تجربة عشوائية. تسمى المجموعة التي تتالف من كل النتائج الممكنة التي تحصل في تجربة عشوائية (فضاء العينة) وكل نتيجة من النتائج تمثل فضاء العينة. مثلاً عند رمي قطعة نقود مرة واحدة فان فضاء العينة (A) سوف يكون اما صورة (A) أو كتابة (T) .

$$A = (\overset{H}{\bullet} \quad \overset{T}{\bullet}) , \quad A = \{H \ , \ T\}$$

اما عند رمي قطعتين من النقود مرة واحدة فان فضاء العينة سوف يتكون من اربعة نتائج ممكنة الحصول والذي يساوي احتمالين مضروب في احتمالين ويساوي اربعة احتمالات.

$$\mathbf{A} = (\overset{HH}{\bullet\bullet} \quad \overset{HT}{\bullet\bullet} \quad \overset{TH}{\bullet\bullet} \quad \overset{TT}{\bullet\bullet}) = \{\mathbf{HH}, \mathbf{HT}, \mathbf{TH}, \mathbf{TT}\}$$

مثال آخر عندما نرمي زهرة النرد مرة واحدة فان فضاء العينة سوف يكون:-

$$A = \{1, \ 2, 3, 4, 5, 6 \ \}$$

حيث ان فضاء العينة يتكون من ستة نتائج ممكنة الحدوث , ولكن لا يمكن اي من الارقام سوف يظهر على السطح العلوي لزهرة النرد الا بعد اجراء التجربة.

2. الحادث Event :

يمثل الحادث نتيجة في أو عدة نتائج في فضاء العينة (عنصر او عدة عناصر) التي ممكن ان تحصل بعد اجراء التجربة. يرمز الى الحادث بالرمز (E) , بمعنى آخر ان الحادث يمثل مجموعة جزئية من فضاء العينة.

مثلاً: عند رمي قطعة نقود مرة واحدة ونحصل على صورة (H) , هذه النتيجة التي حصلنا عليها بعد اجراء التجربة تسمى حادث (Event) وتتكون من نقطة واحدة او عنصر واحد من فضاء العينة الذي هو صورة او كتابة (H or T) , والحادث هو صورة (H) .مثال نخر عند رمي زهرة الطاولة فان فضاء العينة هو :

A = { 1 , 2 , 3 , 4 , 5, 6 } , وان الحصول على عدد زوجي تسمى حادث وهو الذي يمثل ظهور احد الارقام الزوجية { 2 , 4 , 6 } من مجموع فضاء العينة, هذا الحادث اما ان يكون حادث بسيط عندما يتكون من نقطة واحدة في فضاء العينة , أو يكون حادث مركب اذا اشتمل على حاتين أو اكثر من الحالات التي تظهر في فضاء العينة.

نتائج التجارب العشوائية تنقسم الى ثلاثة انواع من وجهة نظر (نظرية الاحتمال) وهي كما يلي:

1. حوادث أو نتائج مؤكدة :

وهي النتائج أو الحوادث التي لا بد من وقوعها أو حدوثها عند اجراء التجربة .

مثال : اذا تم القاء تفاحة في الهواء فاننا نعلم مسبقاً انها لا بد وان تسقط على الارض. هنا التجربة هي القاء التفاحة في الهواء والنتيجة هي سقوط التفاحة على الارض, فانه يقال ان احتمال وقوعها يساوي الواحد.

مثال آخر: اذا كان لدينا صندوق يحتوي على (8) كرات بيضاء اللون وسحنت منه كرة واحدة فلا بد ان تكون الكرة المسحوبة بيضاء. هنا التجربة هي سحب كرة من الصندوق والنتيجة هي الحصول على كرة بيضاء , وبما ان الكرة لا بد ان تكون كرة بيضاء فان النتيجة لا بد ان تكون مؤكدة. اذن اذا كانت الحوادث مؤكدة الوقوع, فانه يقال ان احتمال وقوعها يساوي واحد (1) . في المثال الاول ان احتمال سقوط التفاحة يساوي واحد (1) , وفي المثال الثاني ان احتمال ان تكون الكرة المسحوبة بيضاء اللون يساوي واحد (1) .

2. حوادث أو نتائج مستحيلة:

هي تلك النتائج أو الحوادث التي تعتبر من المستحيل حدوثها.

مثال: هل يمكن سحب كرة حمراء من صندوق لا يحتوي الاّ على كرات بيضاء؟

التجربة هنا هي سحب رة من الصندوق, والنتيجة المطلوبة هي ان تكون الكرة حمراء, اذا فهذه حادثة مستحيلة الحصول على كرة حمراء من صندوق يحتوي على كرات بيضاء فقط , اي ان احتمال سحب كرة حمراء من صندوق لا يحتوي الاّ على كرات بيضاء يساوي صفر.

3. حوادث أو نتائج غير مؤكدة (محتملة او ممكنة):

وهي نتائج التجارب العشوائية التي ذكرناها سابقاً والتي لا نستطيع ان نتنبأ بوقوعها مسبقاً , ولكننا نستطيع ان باستخدام تعريف الاحتمال ان نحسب احتمالية وقوعها, واذا كانت الحالة محتملة فان احتمال وقوعخا ينحصر بين الصفر والواحد. ان لفظ احتمال يعبر عن مدى توقعنا بحدوث شيء معين, هذا التوقع أو التنبؤ أو التخمين قد يكون كبيراً وقد يكون صغيراً, وتبعاً لذلك قد يكون الاحتمال كبيرا وقد يكون صغيرا , وهذا هو الذي يبعث لدينا الرغبة في اجراء المقارنة بين احتمالي حدوث حادثتين لمعرفة ايهما اكبر احتمالية وذلك يتضح في ما يلي:

لو كان لدينا صندوقان يحتوي كل منهما على كرات متشابهة في الحجم والوزن وكل شيء ما عدا اللون , وكان الصندوق الاول يحتوي على(90) كرة بيضاء و (10) كرات سوداء , والصندوق الثاني يحتوي على (10) كرات بيضاء و (90) كرة سوداء , وعندما يتم سحب كرة واحدة عشوائياً من كل صندوق , فأيهما يكون اكثر احتمالاً في الحصول على كرة بيضاء هو من الصندوق الاول أم من الصندوق الثاني؟

بالتفكير المنطقي السليم يمكننا الحكم على ان احتمال الحصول على كرة بيضاء من الصندوق الاول أكبر من احتمال الحصول على كرة بيضاء من الصندوق الثاني, بسبب ان عدد الكرات البيضاء هو الاقل. هذا يوضح ان كل مانعرفه حتى الان هو مجرد مقارنة الاحتمالات ولكن لم نحدد قيمة هذه الاحتمالات بطيقة عددية, هذا هو الذي دفع العلماء الاوائل في هذا المجال الى وضع تعريف أو علاقات رياضية احصائية نتمكن من خلالها قياس الاحتمال بطريقة عددية.

قبل ان نتطرق الى تعريف نظرية الاحتمال بالطريقة العددية سوف نتطرق هنا الى تعريف وشرح بعض القواعد والاسس التي تعتبر نتاجاً لما قام به العلماء الاوائل من دراسات منتظمة في مجال الاحتمالات .

1. الحالات المتماثلة Equally Likely Cases :

هي تلك الحالات التي تكون لها فرص متكافئة ومتماثلة في امكانية حدوثها (لها نفس الفرصة), مثال على ذلك , عند القاء قطعة معدنية وكانت عملية الالقاء غير متحيزة , فان فرصة ظهور الصورة تعادل تماماً فرصة ظهور الكتابة, وهذا ما يطلق عليه بالحالات المتماثلة . كذلك اذا كان لدينا صندوق يحتوي على (50) كرة بيضاء و (50) كرة سوداء ورغبنا في سحب كرة من هذا الصندوق عشوائياً , سنجد ان فرصة الحصول

على كرة بيضاء تعادل تماماً فرصة الحصول على كرة سوداء, وبذلك تعتبر الحصول على اللونان الابيض والاسود حالتين متماثلتين.

2. الحوادث الشاملة Exhaustive Events :

وهي مجموعة الحوادث التي تمثل فضاء العينة والتي لا بد ان يتحقق واحاً منها على الاقل عند اجراء التجربة. مثال على ذلك عند القاء زهرة الطاولة فان الاوجه الستة لزهرة النرد (1,2,3,4,5,6) تعتبر احداث شاملة. كذلك عند القاء قطعة معدنية , يعتبر الوجهان (H , T) حدثين شاملين.

3. الحوادث المتنافية Mutually Exclusive Events :

وهي الحوادث التي يستبعد حدوثهما معاً (اي استحالة حدوث أو حصول حادثين معا مرة واحدة). مثال على ذلك في تجربة القاء زهرة الطاولة , تعتبر الاوجه الستة حوادث متنافية لعدم امكانية حدوث اثنين منهما في آن واحد, وكذلك في تجربة القاء قطعة نقدية , تعتبر الوجهان (الصورة والكتابة) حدثين متنافيين.

4. الحوادث المستقلة Independent Events :

وهي الحوادث التي اذا وقع احدهما فانه يؤثر على وقوع الحوادث الاخرى , أي انه مرتبط مع النتائج الاخرى. مثال على ذلك اذا كان لدينا صندوق يحتوي على كرات بيضاء وسوداء , وعند سحب كرتان على التوالي بدون اعادة الكرة الاولى الى الصندوق , فان نتيجة السحبة الثانية سوف تتاثر بنتيجة السحبة الاولى. في هذه الحالة يعتبر الحادثين غير مستقلين.

5. الحالات الممكنة Possible Casses :

وهي جميع الحوادث التي يمكن ان تظهر في تجربة معينة. مثال على ذلك , عند رمي قطعة نقود فان الحالات الممكنة الوقوع هي حالتين إما تظهر صورة أو كتابة, كذلك عند رمي زهرة الطاولة فان الحالات الممكنة هي ستة حالات وعند رمي زهرتي طاولة فان الحالات الممكنة سوف تكون $\boldsymbol{6\ x\ 6 = 36}$ حالة.

6. الحالات المؤاتية Favorable Cases :

وتسمى كذلك حالات النجاح, وهي الحالات التي يتحقق فيها ظهور الحادث المراد دراسته. مثال على ذلك عند رمي زهرة الطاولة , فان كان المطلوب هو الحصول على عدد زوجي فان الحالات المؤاتية هي التي يتحقق فيها الحصول على عدد زوجي (2 , 4 , 6).

تعريف الاحتمالات

يوجد للاحتمالات تعاريف مختلفة , واهمها تعريفين اثنين فقط وهي التي لا تحتاج الى مفهوم رياضي متقدم لفهمها وهما:

1. **التعريف الكلاسيكي للاحتمالات**
2. **التعريف التجريبي للاحتمالات**

1. التعريف الكلاسيكي للاحتمالات:

نفترض ان حادثاً معينا نرمز له بالمز (E_1) التي يمكن ان ينتج عنها أو يتحقق في (m) من الحالات (أو عدد مرات اجراء التجربة) وتسمى الحالات الشاملة , المتنافية والمتماثلة وتسمى كذلك الحالات المؤاتية من مجموع (N) من الحالات الممكنة (فضاء العينة) على افتراض ان جميع الحالات الممكنة هذه متساوية الحدوث (Equally Likely Cases) , فان درجة احتمال ظهور الحادث (E_1) والذي يرمز له بالرمز $P(E_1)$ يمكن ان يعرف بالصيغة التالية:

$$P(E_1) = \frac{m}{N} = \frac{\text{عدد الحالات المؤاتية (النجاح)}}{\text{عدد الحالات الممكنة (الكلية)}} = \frac{\text{عدد مرات ظهور الحادث}}{\text{فضاء العينة}} , \quad m \le N \text{ بشرط ان}$$

بالمقابل فان عدم ظهور الحادث (تسمى حالات الفشل) والتي يرمز لها بالرمز $\boldsymbol{P(\overline{E_1})}$ سوف يكون أو يعرف بالصيغة التالية :

$$\boldsymbol{P(\overline{E_1}) = 1 - P(E_1)}$$

ويسمى كذلك عدم ظهور الحادث $\boldsymbol{P(\overline{E_1})}$ بالحادث العكسي, أي ان :

$$\boldsymbol{P(\overline{E_1}) + P(E_1) = 1}$$

مثال (1): اذا كانت التجربة هي القاء زهرة الطاولة فان الحالات الممكنة بهذه الحالة تكون ستة حالات شاملة ومتباينة ومتماثلة , فاذا كان الحادث هو الحصول على عدد زوجي من النقاط , فان الحالات المؤاتية لهذا الحادث هو ثلاث حالات وهي الاعداد (2 , 4 , 6) وهي التي تمثل الاوجه التي تمتلك عدد زوجي من النقاط , وبهذا يكون احتمال وقوع هذا الحادث مساوياً الى :

$$\boldsymbol{\frac{3}{6}} = \frac{\text{عدد الحالات المؤاتية}}{\text{عدد الحالات الممكنة}} = \boldsymbol{\frac{1}{2}}$$

مثال (2): عند القاء قطعتي نقود مرة واحدة (أو القاء قطعة نقود مرتين متتاليتين) , ما هو احتمال الحصول على صورتين؟

الحل: التجربة هي القاء قطعتي نقود.

الحالات الممكنة هي: $(N) = 2\ x\ 2 = 4$ حالات, وذلك لان لكل قطعة لها وجهان ولكل وجه منهما يمكن ان يناظر وجهان في القطعة الثانية وهذه الحالات الاربعة يمكن حصرها بفضاء العينة الذي يمثل:

T T , H T , T H , H H

الحدث هو الحصول على صورتين:

اذن الحالات المؤاتية (H) هي حالة واحدة للحصول على صورتين المتمثلة بظهور (H H) , اذن:

$$\therefore\ P\,(E_1) = \frac{m}{N} = \frac{1}{4}$$

مثال (3): عند القاء زهرة الطاولة مرة واحدة , ما هو احتمال ان يكون مجموع النقاط على السطح العلوي اكثر من اربعة؟

الحل: التجربة هي القاء زهرة طاولة مرة واحدة.

الحالات الممكنة: هي ستة حالات متماثلة وممكنة وشاملة

الحادث: هو ان يكون مجموع النقاط على السطح العلوي لزهرة الطاولة اكبر من اربعة.

اذن : الحالات المؤاتية هي $(n = 2)$ (وهي الحالتين عند ظهور الارقام 5 و 6).

$$\therefore\ P\,(E_1) = \frac{m}{N} = \frac{2}{6} = \frac{1}{3}$$

مثال (4): عند القاء زهرتين من زهرات الطاولة مرة واحدة (أو القاء زهرة طاولة واحدة مرتين), ما هو احتمال ان يكون مجموع النقاط على السطحين العلويين: أ- مساوياً الى (9) , ب- اكثر من (9)

الحل: التجربة : القاء زهرتي طاولة مرة واحدة

عدد الحالات الممكنة أو المتماثلة (N) يساوي (فضاء العينة) , وهي ستة حالات للزهرة الاولى يقابلها ستة حالات للزهرة الثانية , وبذلك سوف يكون عدد الحالات الممكنة مساوياً الى : $\left(6\ x\ 6 = 36\right.$ حالة$\left.\right)$, وكما في الشكل التالي:

6 , 1	5 , 1	4 , 1	3 , 1	2 , 1	1 , 1
6 , 2	5 , 2	4 , 2	3 , 2	2 , 2	1 , 2
6 , 3	5 , 3	4 , 3	3 , 3	2 , 3	1 , 3
6 , 4	5 , 4	4 , 4	3 , 4	2 , 4	1 , 4
6 , 5	5 , 5	4 , 5	3 , 5	2 , 5	1 , 5
6 , 6	5 , 6	4 , 6	3 , 6	2 , 6	1 , 6

الحالات المذكورة في الجدول تمثل (36) نتيجة ممكنة أو تمثل فضاء العينة التي من الممكن ان تتحقق اي منها. مثلاً النتيجة (5 , 2) , معناها ان الزهرة الاولى نتيجتها الوجه الذي عليه (5) نقاط والزهرة الثانية الوجه الذي عليه نقطتين.

اذن الحادث: للحالة الاولى (أ), هو ان يكون مجموع النقاط على السطحين العلويين يساوي (9).

الحالات المؤاتية: هي تلك الحالات التي نحصل فيها على الرقم الذي يساوي بمجموعه (9) , وهي تلك الحالات المؤشرة في المربعات في الجدول اعلاه والمحاطة بالخط السميك وعددها (4) اربعة حالات, اي ان $(\boldsymbol{n = 4})$.

$$\therefore \quad \boldsymbol{P\,(E_1) = \frac{n}{N} = \frac{4}{36} = \frac{1}{9}}$$

الحالة الثانية (ب) , التي يكون مجموع النقاط على السطحين العلويين اكثر من (9).

اذن الحالات المؤاتية : هي كل الحالات للمربعات الظاهرة في الجدول التي تمثل مجموع الارقام للزهرتين اكثر من (9) , والتي عددها يساوي ستة مربعات كما مؤشرة في الجدول اعلاه والمحاطة بالخط المزدوج, اي ان $(\boldsymbol{n = 6})$.

$$\therefore \quad \boldsymbol{P\,(E_1) = \frac{n}{N} = \frac{6}{36} = \frac{1}{6}}$$

<u>مثال (5)</u>: صندوق يحتوي على ثلاث كرات بيضاء وخمس كرات سوداء, فاذا تم سحب كرة واحدة من هذا الصندوق عشوائياً , ما هو احتمال ان تكون هذه الكرة سوداء؟

الحل: ان فضاء العينة لهذه التجربة هة ثمانية نتائج , اي ان عدد الحالات الممكنة يساوي ثمانية (8).

A = فضاء العينة ● ● ● ● ● ○ ○ ○

الحالات المؤاتية : وهي عدد حالات النجاح التي من الممكن ان نحصل على كرة سوداء وهي خمسة حالات او خمسة طرق (بسبب ان عدد الكرات السوداء هي (5) كرات) , لذلك فان احتمال ان تكون الكرة سوداء هي:

$$\therefore \; P\,(E_1) = \frac{n}{N} = \frac{5}{8}$$

مثال (6): صندوق يحتوي على (6) كرات حمراء و(4) كرات بيضاء و(5) كرات صفراءاذا سحبت كرة واحدة عشوائياً , ما هي درجة ان تكون هذه الكرة:

a) كرة حمراء

b) كرة غير حمراء

الحل:

a) عدد الحالات الناجحة هي التي تكون فيه الكرة حمراء هي (6) حالات . عدد الحالات الممكنة لوقوع الحادث يساوي (15) , وهو عدد الكرات الكلية في الصندوق.

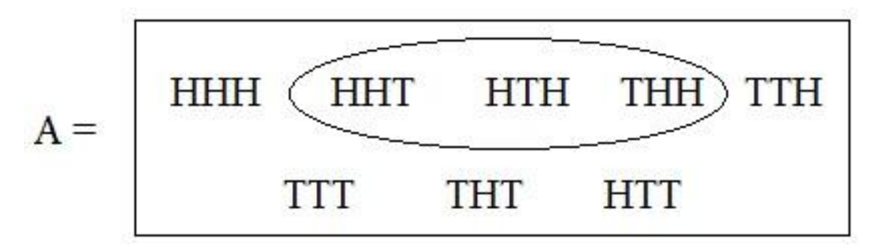
$$\therefore \; P\,(E_1) = \frac{n}{N} = \frac{6}{15}$$

b) عدد الحالات الناجحة التي تكون فيه الكرة غير حمراء هي (4)كرات بيضاء + (5)كرات صفراء = (9) كرات غير حمراء.

$$\therefore \; P\,(E_1) = \frac{n}{N} = \frac{9}{15}$$

مثال (7): رميت قطعة نقود ثلاث مرات , فما هو احتمال الحصول على الصورة مرتين.

الحل: في كل رمية لقطعة النقود هنالك احتمالين , إما نحصل على صورة (H) أو كتابة (T) , بما ان هناك ثلاث رميات لقطعة النقود , وعليه فان فضاء العينة يساوي $(2x2x2 = 8)$. بمعنى آخر ان فضاء العينة يمكن تمثيله بالشكل التالي:



بهذه الحالة فان عدد الحالات الممكنة التي نحصل فيها على الصورة مرتين من خلال اجراء تجربة رمي قطعة النقود ثلاث مرات هي:

$$\therefore \ \boldsymbol{P\,(E_1) = P(2H) = \frac{n}{N} = \frac{3}{8}}$$

<u>مثال(8):</u> في مصنع للمصابيح الكهربائية , نحصل من كل (1000) مصباح منتج نحصل على (50) مصباح ردئ, اخترنا احد المصابيح من انتاج المصنع:

A. ما هو احتمال الحصول على مصباح جيد؟

B. ما هو احتمال الحصول على مصباح ردئ؟

<u>الحل:</u> انتاج المصنع هو العدد الكلي للحالات , اي ان N=1000,

A. ان احتمال الحصول على مصباح جيد :

الحادث $P(E_1)$ هو الحصول على مصباح جيد , اي ان m=950

$$P(E1) = \frac{m}{N} = \frac{950}{1000} = 0.95$$

B. ان احتمال الحصول على مصباح ردئ :

الحادث $P(E_1)$ هو الحصول على مصباح ردئ , اي ان m=50

$$P(E1) = \frac{m}{N} = \frac{50}{1000} = 0.05$$

التعريف النسبي للاحتمال:

في حالة اذا تم اجراء تجربة عشوائية , ثم اعيدت هذه التجربة لعدد (n) من المرات , فان احتمال حدوث أو تحقق (تحقق حصول) الحادث (E_1) , ممكن ان يعبر عنه بصيغة النسبة المئوية أو نسبة حصول الحادث , ويمكن ان يعرف بالعلاقة التالية:

$$P\ (E_1) = \frac{n}{N}\ x\ 100\ \%$$

مثال : تم القاء زهرة الطاولة (100) مرة , فكان عدد مرات ظهور السطح الذي عليه ثلاث (3) نقاط هو (20) مرة . إحسب نسبة ظهور الرقم (3) في هذه التجربة؟

الحل: احتمال ظهور الوجه الذي عليه (3) نقاط هو:

$$P\ (E_1) = P(3) = \frac{n}{N}\ x\ 100\ \% = \frac{20}{100}\ x100\ \% = 20\%$$

نلاحظ من هذه التجربة ان الاحتمال يحسب بعد اجراء التجربة ومشاهدة نتائجها , هذه الاحتمالات تسمى (Posteriori Probability) , اما الاحتمالات عن توقع حدوث الحادث تسمى قبليةPoriori Probability .

بعض خواص الاحتمال:

الخاصية الاولى:

اذا رمزنا لاحتمال حصول الحادث بالرمز $P(E_1)$, ونرمز لاحتمال عدم حصول الحادث بالرمز $P(\bar{E}_1)$

فان: $P(E_1) + P(\bar{E}_1) = 1$, والسبب في ذلك ان :

$P\ (E_1) = \frac{n}{N}$, وكذلك $P(\bar{E}_1) = \frac{N-n}{N}$, وبهذه الحالة فان : $P(E_1) + P(\bar{E}_1) = 1$

مثال : صندوق يحتوي على (6) كرات حمراء و(4) كرات بيضاء , و (5) كرات صفراء , فاذا سحبت منه كرة واحدة عشوائياً , فما هي درجة احتمال ان تكون هذه الكرة : أ- حمراء , ب- غير حمراء

الحل:

أ- درجة احتمال ان تكون الكرة حمراء هي : $P\ (E_1) = P\ (Red) = \frac{n}{N} = \frac{6}{15}$

ب- درجة احتمال ان تكون الكرة غير حمراء هي:

$$P(\bar{E}_1) = P(White + Yellow) = \frac{N-n}{N} = \frac{15-6}{15} = \frac{9}{15}$$

$$\therefore\ P\ (E_1) +\ P(\bar{E}_1) = \frac{6}{15} + \frac{9}{15} = \frac{15}{15} = 1$$

الخاصية الثانية:

إن درجة احتمال حصول اي حادث تتراوح بين الصفر والواحد, أي ان :

$$0 \le P\,(E_1) \le 1$$

اذا كانت درجة احتمال حصول الحادث $(E_1) = 1$, سمي الحادث بهذه الحالة (حادث أكيد) , اي ان:

$$P\,(E_1) = 1 \quad \rightarrow \quad \text{عندما يكون } n = N$$

اما اذا كانت درجة احتمال حصول الحادث $(E_1) = 0$, سمي الحادث بهذه الحالة (حادث مستحيل), اي ان:

$$P\,(E_1) = 0 \quad \rightarrow \quad \text{عندما يكون } n = 0$$

مثال : صندوق يحتوي على (20) كرة بيضاء , فادا سحبت منه كرة واحدة عشوائياً , فما هو احتمال :

أ- ان تكون الكرة بيضاء , ب- ان تكون الكرة سوداء

الحل:

أ- احتمال ان تكون الكرة بيضاء : يسمى حادث اكيد , $P(white) = \frac{n}{N} = \frac{20}{20} = 1$

ب- احتمال ان تكون الكرة سوداء : يسمى حادث مستحيل, $P(black) = \frac{n}{N} = \frac{0}{20} = 0$

الفصل التاسع

التباديل Permutation

تعرف التباديل بانها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة اشياء باخذها كلها او بعضها. ونعني بالاشياء هي (عدد النماذج , عدد العناصر , عدد المجاميع , عدد القيم , ... الخ). أو بمعنى آخر هي عدد طرق الاختيار الممكنة لمجموعة من الاشياء مع الاخذ بنظر الاعتبار الترتيب , (حيث ان الترتيب مؤثر ومهم) . يرمز للتباديل بالرمز (nPr) أو P_r^n أو $\binom{n}{r}$, وتقرأ تباديل r من n , حيث ان :

عدد طرق الاختيار = r

عدد المجاميع أو الاشياء أو النماذج الكلية = n

مثال: اذا كان لدينا مجموعة شاملة هي : $M = \{1 , 2 , 3\}$, مكونة من ثلاث عناصر أو اشياء أو نماذج , ناخذ مجموعة جزئية (m) من المجموعة الشاملة (M) , والتي تتكون من عنصرين وهي :

$$m = \{1 , 2\} \ , \ m = \{2 , 1\} \ , \quad m = \{2 , 3\} \ , m = \{3 , 2\}$$

نلاحظ ان المجموعة الجزئية $m = \{1 , 2\}$ هي ليست نفسها المجموعة الجزئية $m = \{2 , 1\}$, لان الترتيب مختلف بالعناصر , حيث ان الترتيب مهم في حالة التباديل.

هناك عدة قوانين للتباديل تعتمد على طبيعة طرق الاختيار وهي:

A. في حالة اذا كانت عدد طرق الاختيار اقل من عدد النماذج $[n > r]$, فان الصيغة الرياضية للتباديل في هذه الحالة سوف تكون :

$$P_r^n = nPr = \frac{n!}{(n-r)!}$$

مثال (1): احسب P_4^{10} , حيث ان :

عدد طرق الاختيار = r = 4 , عدد النماذج = n = 10

الحل:

$$P_4^{10} = nPr = 10P4 = \frac{10\,!}{(10-4)\,!} = \frac{10\,!}{6\,!} = \frac{10x9x8x7x6x5x4x3x2x1}{6x5x4x3x2x1}$$
$$= 5040$$

مثال (2): اذا كان لدينا اربعة حروف هي (A , B , C , D) , يراد اختيار حرفان فقط من هذه الحروف. ما هي عدد طرق الاختيار التي يمكن بها اختيار هذين الحرفين؟

الحل:

$$P_2^4 = nPr = 4P2 = \frac{4\,!}{(4-2)\,!} = \frac{4\,!}{2\,!} = \frac{4x3x2x1}{2x1} = 12$$

بطريق اخرى يمكن توضيح عدد طرق الاختيار بهذه الطريقة المطولة :

(A,B,C,D) = {AB, AC, AD, BC, BD, CD, BA, CA, DA, CB, DB, DC}

B. في حالة اا1 كانت عدد طرق الاختيار تساوي عدد النماذج الكلية , اي ان : $[n = r]$, بهذه الحالة يكون : $nPr = nPn = n! = nx(n-1)\text{x}(n-2)\text{x}(n-3)\text{x}\ldots\ldots\ldots\text{etc.}$.

مثال (1): اذا اراد طالب ان يرتب اربعة كتب مختلفة المواضيع على رف مكتبة , بكم طريقة ممكن ان يرتب هذه الكتب؟

الحل: في هذه الحالة نلاحظ ان عدد النماذج (الكتب) تساوي عدد طرق الاختيار. لذلك نطبق القانون التالي:

$nPn = n!$ حيث ان $n = r$

$4P4 = 4! = 4x(4-1)x(4-2)x(4-3) = 4x3x2x1 = 24$ طريقة ترتيب الكتب

مثال (3): كتبت الارقام من (1) الى (9) على بطاقات ووضعت في صندوق , ثم سحبت منه (5) بطاقات (الواحدة بعد الاخرى) , كم عدد خماسي بارقام مختلفة يمكن تكوينه من البطاقات المسحوبة اذا كان :

أ- $(n > r)$ ب- $(n = r)$

الحل:

أ- في حالة $(n > r)$, نطبق العلاقة الرياضية التالية: $nPr = \frac{n!}{(n-r)!}$

ان المرتبة الاولى من الرقم الخماسي يمكن اختياره بتسعة طرق , وبعد اختيار الرقم الاول , تبقى ثمانية ارقام نختار احدهما للمرتبة الثانية , أي يمكن اختيار المرتبة الثانية بثمانية طرق , ويتم اختيار المرتبة الثالثة بسبعة طرق , وهكذا الى المرتبة الخامسة , بخمسة طرق.

$$nPr = 9P9 = \frac{9!}{(9-5)!} = \frac{9!}{4!} = \frac{9x8x7x6x5x4x3x2x1}{4x3x2x1} = 15120$$

اي ان هناك (15120) طريقة ترتيب لتكوين رقم خماسي مكون من خمسة ارقام.

ب- في حالة $(n > r)$, نطبق العلاقة الرياضية التالية: $n = r$.

$$9P9 = n = r = 9! = 9x8x7x6x5x4x3x2x1 = 362880$$

اي ان هنلك (362880) ترتيب ممكن تكوينه من تسعة ارقام باخذها كلها.

<u>التباديل في حالة وجود مجموعات متشابهة:</u>

ندرس الان عدد الطرق المختلفة التي يمكن بها ترتيب حروف كلمة (باب) , وقد تبدو للوهلة الاولى ان عدد الطرق هو (6) ناتج من $3! = 3x2x1 = 6$, الا ان تكرار الحرف (ب) مرتين يمنع استعمال او لا يمكن استخدام هذه القاعدة لعدم امكانية التمييز بين (ب) في اول الكلمة و(ب) في آخر الكلمة , اي ان عدد الحروف هي (3) أثنان منها متشابهة وبذلك يصبح عدد الممكن ترتيب حروف كلمة (باب) هي:

$$\frac{3!}{2!\,1!} = \frac{3x2x1}{2x1x1} = 3$$

وبذلك في حالة وجود مجاميع او اشباء او نماذج او عناصر متشابهه في هذه الحالة نطبق القانون التالي:

$$nPm = \frac{n!}{m1!\,x\,m2!\,x\,m3!\,x\,\ldots}$$

حيث ان :

n = عدد المجاميع او النماذج او الاشياء

m = عدد التكرار لكل نموذج او مجموعة

مثال (3): ما هو عدد التباديل التي يمكن تكوينها من احرف كلمة (statistics) .

الحل : ان مجموع احرف كلمة statistics هي (10) احرف , اي ان (n=10) .

الحرف (s) مكرر ثلاث مرات = $m_1 = 3$

الحرف (t) مكرر ثلاث مرات = $m_2 = 3$

الحرف (a) مكرر مرة واحدة = $m_3 = 1$

الحرف (i) مكرر مرتان = $m_4 = 2$

الحرف (c) مكرر مرة واحدة = $m_5 = 1$

نطبق صيغة القانون الخاص بالمجاميع المتشابهة على هذه الحالة وكما يلي :

$$\frac{n!}{m1!\,xm2!\,xm3!\,xm4!\,xm5!} = \frac{10!}{3!\,x3!\,x1!\,x2!\,x1!}$$
$$= \frac{10x9x8x7x6x5x4x3x2x1}{(3x2x1)x(3x2x1)x(1)x(2x1)x(1)} = 50400$$

مثال (4): بكم طريقة ممكن ترتيب خمس قطع ملونة على خط مستقيم ؟

الحل: كما في الشكل التالي يمكن ان نوضح طريقة الحل:

اللون 1	اللون 2	اللون 3	اللون 4	اللون 5

الموقع الاول يمكن ان نضع فيه اية قطعة ملونة من القطع الخمسة

الموقع الثاني يمكن ان فيه اية قطعة ملونة من القطع الاربعة الملونة الباقية

الموقع الثالث يمكن ان فيه اية قطعة ملونة من القطع الملونة الثلاثة الباقية

الموقع الرابع يمكن ان فيه اية قطعة ملونة من القطع الملونة الاثنان الباقية

الموقع الخامس يمكن ان نضع فيه القطعة الملونة الاخيرة

ثم نطبق القانون التالي: $\boldsymbol{nPn = n!}$

$$\boldsymbol{nPn = n! = 5x4x3x2x1 = 120}$$

اي ان لدينا 120 طريقة اختيار لترتيب القطع الملونة

مثال (3): بكم طريقة يمكن لمجموعة من سبعة اشخاص ان يجلسوا على سبعة مقاعد؟

الحل:

$$\boldsymbol{nPn = n! = 7! = 7x6x5x4x3x2x1 = 5040} \text{ طريقة}$$

مثال (4): لدينا خمس قطع حمر , قطعتان بيضاء , وثلاث قطع زرقاء, رتبت على خط مستقيم , بكم طريقة يمكن ان نرتب هذه القطع؟

الحل: نطبق قانون المجموعات المتشابهة:

$$\boldsymbol{nPm = \frac{n!}{m1!\,x\,m2!\,x\,m3!} = \frac{10!}{5!\,x2!\,x3!} = 2520} \text{ طريقة ترتيب}$$

مثال (5): ما هي عدد طرق الاختيار التي يمكن تكوينها من ثلاثة ارقام مأخوذة من بين الارقام (3 و4 و5 و6 و7 و8) في حالة:

أ- دون تكرار الرقم في العدد ب- يمكن تكرار الرقم في العدد

الحل: أ- نستخدم القانون التالي في حالة عدم السماح بتكرار الرقم في العدد وهو:

$$\boldsymbol{P_3^6 = 6\ x\ 5\ x\ 4 = 120} \text{ عدد}$$

ب- في حالة السماح بتكرار العدد في الرقم :

تكون عدد طرق اختيار رقم الآحاد = 6

تكون عدد طرق اختيار رقم العشرات = 6

تكون عدد طرق اختيار رقم المئات = 6

وبموجب هذا المبدأ تكون عدد طرق الاختيار = طريقة $\mathbf{216 = 6\ x\ 6\ x\ 6}$

مثال (6): بكم طريقة ممكن لعشرة طلاب ان يجلسوا على اربعة مقاعد فقط؟

الحل:

1. الكرسي الاول ممكن ان يجلس عليه اي شخص من الاشخاص العشرة (اختيار واحد من عشرة).
2. الكرسي الثاني ممكن ان يجلس اي شخص من الاشخاص التسعة الباقين (اختيار واحد من تسعة).
3. الكرسي الثالث ممكن ان يجلس عليه اي شخص من الاشخاص الثمانية الباقين (اختيار واحد من ثمانية)
4. الكرسي الرابع ممكن ان يجلس عليه اي شخص من الاشخاص السبعة الباقين (اختيار واحد من سبعة)

اي نختار اربعة اشخاص من بين عشرة اشخاص للجلوس على ارعة كراسي, نطبق الصيغة الرياضية التالية :

طريقة $\boldsymbol{nPn = nPr = n! = 10x9x8x7 = 5040}$

مثال (7): جد عدد التباديل للحروف (أ , ب , ج) الماخوذة منها اثنين في كل مرة؟

الحل: يمكن حل السؤال بطريقتين:

1. $\boldsymbol{nPr = \frac{n!}{(n-r)} = \frac{3!}{(3-2)!} = \frac{3x2x1}{1x1} = 6}$
2. $\boldsymbol{nPr = 3P2 = 3x2\ 6}$

الفصل العاشر

التوافيق Combination

يقصد بالتوافيق هو ان كل مجموعة تتكون من كل أو من بعض الاشياء أو النماذج أو العناصر , وعند اختيار بعض من هذه النماذج أو العناصر , فان طرق الاختيار تتم بغض النظر عن ترتيب النماذج أو العتاصر داخل هذه المجموعة , عندئذ تسمى هذه العملية بالتوافيق. إن التوافيق لا تهتم بالترتيب وتعطي الاهمية فقط لمجموعة الاشياء التي نختارها. لذلك يمكن تعريف التوافيق بانها (عدد طرق الاختيار غير المرتب (العشوائي) التي يمكن تكوينها من عدة اشياء باخذها كاها او بعضها, (اي ان الترتيب غير مهم في حالة التوافيق).

يرمز للتوافيق بالرمز : C_r^n أو $\binom{n}{r}$ $nCr\ H,$

حيث إن:

عدد طرق الاختيار = r

عدد النماذج أو عدد المجاميع أو عدد الاشياء = n

بمعنى آخر ان التوافيق هي عبارة عن عدد طرق الاختيار الممكنة لمجموعة من الاشياء دون النظر لاعتبار الترتيب. الصبغة الرياضية للتوافيق هي:

$$nCr = \binom{n}{r} = C_r^n = \frac{n!}{r!\,(n-r)!} = \frac{nPr}{r!}$$

مثال (1): اذا كان لدينا مجموعة شاملة هي $M = \{1, 2, 3\}$ مكونة من ثلاث عناصر أو نماذج هي 3 و 2 و 1 نلاحظ ان المجموعة الجزئية m من المجموعة الشاملة $\mathbf{M}$: تتكون من عنصرين هما

$m = \{1, 2\}, m = \{1, 3\}$, $m = \{2, 3\}$, $m = \{2, 1\}$, $m = \{3, 2\}$, $m = \{3, 1\}$

, وان $m = \{2, 1\}$ هي نفسها $m = \{1, 2\}$ نلاحظ ان المجموعة الجزئية $m = \{3, 1\}$ هي نفسها $m = \{1, 3\}$, ... وهكذا, اي ان اختيار العناصر ضمن المجموعة المجموعات الجزئية تتم على اساس اختيار عنصرين من بين ثلاث عناصر بصورة عشوائية دون الاهتمام بترتيب تلك العناصر. عملية الاختيار هذه تسمى توافيق. في هذا المثال تسمى توافيق (3) ماخوذة اثنين اثنين وصيغتها الرياضية كما يلي:

$$3C2 = \frac{3!}{2!\,(3-2)!}$$

مثال(2): كم لجنة ثلاثية يمكن تكوينها من ستة اشخاص؟

الحل:

$$6C3 = \frac{6!}{3!(6-3)!} = \frac{6!}{3!x3!} = \frac{6x5x4x3x2x1}{3x2x1x3x2x1} = \frac{120}{6} = 20$$

مثال (3): أوجد توافيق 7C4

الحل:

$$7C4 = \frac{7!}{4!\,(7-4)!} = \frac{7!}{4!\,x\,3!} = \frac{7x6x5x4x3x2x1}{4x3x2x1x3x2x1} = \frac{310}{6} = 35$$

مثال (4): أوجد توافيق 4C4

الحل:

$$4C4 = \frac{4!}{4!\,(4-4)!} = \frac{4!}{4!\,x\,0!} = \frac{4x3x2x1}{4x3x2x1x1} = \frac{24}{24} = 1\,, where\; 0! = 1$$

مثال (5): ما عدد طرق الاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من خمسة اشخاص من مجموع تسعة اشخاص؟

الحل:

$$9C5 = \frac{9!}{5!\,(9-5)!} = \frac{9!}{5!\,x\,4!} = \frac{9x8x7x6x5x4x3x2x1}{5x4x3x2x1x4x3x2x1} = \frac{3024}{24} = 126$$

هناك قاعدتان اساسيتان يعتمد عليها كل من التباديل والتوافيق وهي:

القاعدة الاولى:

عند اجراء تجربة لغرض الحصول على نتيجة معينة , تسمى هذه النتيجة (بالحادث)ويرمز له بالرمز (E). أما اذا كان لدينا حادثين أو نتيجتين عند اجراء التجربة , في هذه الحالة يرمز للحادث الاول بالرمز (E_1) والحادث الثاني بالرمز (E_2) .

نرمز الى عدد الطرق الممكنة للحصول على الحادث (E_1) بالرمز (n). ونرمز الى عدد الطرق الممكنة للحصول على الحادث (E_2) بالرمز (m) . في هذه الحالة اذا كان هذان الحادثان E_1 و E_2 , مستقلان , اي لا يوجد اي تاثير من وقوع الحادث E_1 على نتيجة وقوع الحادث E_2 , عند ذلك يكون مجموع عدد الطرق الممكنة (أي التوافيق) لوقوع الحادثان معاً E_1 و E_2 هو :

عدد الطرق $= m\,x\,n$ ← في حالة الحوادث المستقلة

مثال : بكم طريقة يمكن اختيار لجنة مكونة من ثلاث رجال وسيدتين من بين سبعة رجال وخمس سيدات ؟

الحل:

يمكن اختيار ثلاثة رجال من بين سبعة رجال بعدد طرق عددها C_3^7 , ويمكن اختيار سيدتين من بين خمسة سيدات بعدد طرق عددها C_2^5 . اذن عدد طرق الاختيار الى اللجنة سوف يكون:

$$\text{عدد طرق الاختيار} = m\,x\,n = C_3^7 \text{ x } C_2^5 = \frac{7!}{3!\,x(7-3)!} \text{ x } \frac{5!}{2!\,(5-2)!} = 350$$

القاعدة الثانية:

اذا كان عدد الطرق الممكنة لوقوع الحادث (E_1) هي (n) طريقة , وان عدد الطرق الممكنة لوقوع الحادث (E_2) هي (m) طريقة , وان الحادثين E_1 و E_2 هما حادثين (متنافيان), بهذه الحالة فان عدد الطرق الممكنة لوقوع الحادثين E_1 و E_2 تعرف بالعلاقة التالية:

$$\text{عدد الطرق الممكنة} = \mathbf{m} + n$$

مثال(1):

صندوق فيه عشرة كرات حمراء وستة كرات بيضاء , سحبت منه اربعة كرات معاً بدون ارجاع. ما عدد الطرق الممكنة التي تكون فيه الكرات المسحوبة من نفس اللون؟

الحل:

$$C_4^{10} + C_4^6 = \frac{10!}{4!\,(10-4)!} + \frac{6!}{4!\,(6-4)!} = 225$$

مثال(2): لدى شخص خمسة قطع من النقود من مختلف الفئات, كم مجموع مختلف من النقود نستطيع ان نكون من هذه القطع الخمسة.

الحل:

ان عدد طرق الاختيار الممكنة هي ان نختار :

1. نختار قطعة واحدة من مجموع خمس قطع
2. نختار قطعتين من مجموع خمس قطع
3. نختار ثلاث قطع من مجموع خمس قطع
4. نختار اربعة قطع من مجموع خمس قطع
5. نختار خمس قطع من مجموع خمس قطع

اذن مجموع عدد طرق الاختيار هي : $C_1^5 + C_2^5 + C_3^5 + C_4^5 + C_5^5$

$$= \frac{5!}{1!(5-1)!} + \frac{5!}{2!(5-2)!} + \frac{5!}{3!(5-3)!} + \frac{5!}{4!(5-4)!} + \frac{5!}{5!(5-5)!} = 31$$

مثال (3): أريد تشكيل لجنة مكونة من عالمي رياضيات اثنان , وثلاث فيزيائيين , من اصل خمس علماء رياضيات وسبعة فيزيائيين . بكم طريقة يمكن القيام بذلك اذا كان :

1. بالامكان ضم اي عالم رياضيات او فيزيائي.
2. يجب ان يوجد عالم فيزيائي في اللجنة.
3. يجب ان لا يتواجد عالم الرياضيات في اللجنة.

الحل:

1. يمكن ان يتم انتقاء عالمي الرياضيات الاثنان من اصل خمس علماء بعدد طرق قدرها $5C2$, ويتم انتقاء ثلاث علماء فيزيائيين من اصل سبعة علماء بعدد طرق قدرها $7C3$, وبذلك يصبح عدد طرق الاختيار الكلية تساوي $5C2 \ x \ 7C3$.

$$\text{عدد الطرق} = 5C2 \ x \ 7C3 = \frac{5!}{2!(5-2)!} \ x \ \frac{7!}{3!(7-3)!} = \frac{5!}{2!x3!} \ x \ \frac{7!}{3!x4!}$$
$$= \frac{5x4x3x2x1}{2x1x3x2x1} \ x \ \frac{7x6x5x4x3x2x1}{3x2x1x4x3x2x1} = \frac{20}{2} \ x \frac{210}{6} = 350$$

2. يتم اختيار عالمي الرياضيات الاثنان من اصل خمسة علماء بعدد طرق قدرها $5C2$, أما عالمي الفيزياء , باعتبار ان عالم فيزيائي واحد موجود اصلا في اللجنة , عندئذ يبقى (6) علماء وبذلك تصبح عدد طرق الاختيار عالمي فيزياء اثنان من اصل (6) علماء هي $6C2$, اذن عدد طرق الاختيار الكلية تكون $5C2$ x $6C2$.

$$\text{عدد الطرق} = 5C2 \ x \ 6C2 = \frac{5!}{2!(5-2)!} \ x \ \frac{6!}{2!(6-2)!} = \frac{5!}{2!x3!} \ x \ \frac{6!}{2!x3!}$$
$$= \frac{5x4x3x2x1}{2x1x3x2x1} \ x \ \frac{6x5x4x3x2x1}{2x1x3x2x1} = \frac{20}{2} \ x \frac{30}{2} = 150$$

3. في حالة استبعاد عالمي الرياضيات من اللجنة ويبقى ثلاث علماء , عندئذ يجب اختيار عالمي رياضيات اثنان منهم الى اللجنة من اضل ثلاث علماء بعدد طرق $3C2$. يمكن ان اختيار ثلاث علماء فيزيائيين من اصل سبعة علماء بعدد طرق قدرها $7C2$, وبذلك يصبح عدد الطرق الكلية للاختيار هي $3C2 \ x \ 7C2$.

$$\text{عدد الطرق} = 3C2\ x\ 7C3 = \frac{3!}{2!\,(3-2)!}\ x\ \frac{7!}{3!\,(7-3)!} = \frac{3!}{2!\,x1!}\ x\ \frac{7!}{3!\,x4!}$$

$$= \frac{3x2x1}{2x1x1}\ x\ \frac{7x6x5x4x3x2x1}{3x2x1x4x3x2x1} = \frac{3}{1}\ x\frac{210}{6} = 105$$

مثال (4): صندوق يحتوي على (6) كرات حمراء و (4) كرات سوداء و (2) بيضاء , بكم طريقة يمكن اختيار (5) كرات بحيث تكون (3) حمراء و (2) سوداء.

الحل:

بما ان الحوادث مستقلة:

1. عدد طرق اختيار (3) كرات حمراء هي $6C3$.
2. عدد طرق اختيار (2) كرات سوداء هي $4C2$.
3. مجموع عدد طرق الاختيار سوف تكون $n\ x\ m$

$$n\ x\ m = 6C3 \text{ x } 4C2 = \frac{6!}{3!\,(6-3)!} \text{ x } \frac{4!}{2!\,(4-2)!} = 120$$

مثال (5): بكم طريقة يمكن اختيار مجموعة من النماذج موجودة في المخزن مكونة من ثلاثة نماذج رمال ونموذجين من الاطيان , من بين عشرة نماذج رمال وثمانية نماذج من الاطيان.

الحل:

1. يمكن اختيار ثلاثة نماذج من الرمال من بين عشرة نماذج من الرمال بعدد طرق قدرها $10C3$.
2. يمكن اختيار نموذجين من الاطيان من بين ثمانية نماذج من الاطيان بعدد طرق قدرها $8C2$.
3. مجموع عدد طرق الاختيار سوف يكون :

$$10C3\ x\ 8C2 = \frac{10!}{3!\,(10-3)!}\ x\ \frac{8!}{2!\,(8-2)!} = \frac{720}{6}\ x\frac{56}{2} = 10080$$

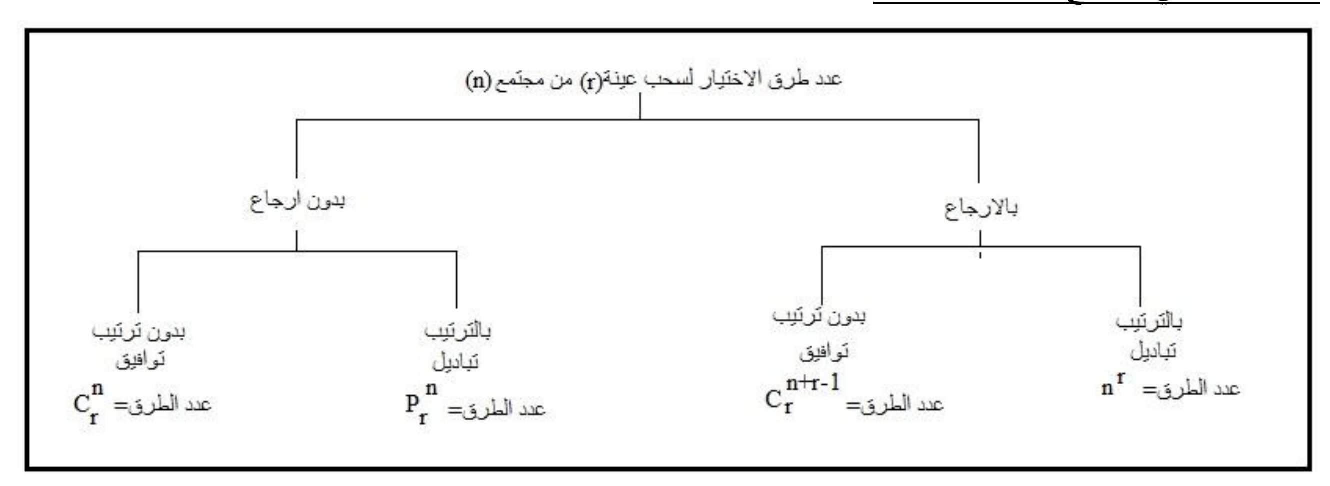
<u>تجارب سحب عينة من مجتمع (Population) :</u>

عند سحب عينة او نموذج من مجموعة نماذج أو (مجتمع) وهي التي تمثل الظاهرة المدروسة , بهذه الحالة يجب مراعاة ما يلي :

1. السحب بالارجاع يعني ان كل عينة تسحب تعاد الى المجموعة الاصلية قبل البدأ أو الشروع بسحب عينة اخرى.

2. السحب بدون ارجاع يعني ان العينة التي تسحب لا تعاد مرة اخرى الى المجموعة الاصلية.

المخطط التالي يوضح عمليات السحب



ملاحظة : اذا لم تذكر طريقة السحب فتعتبر دون ارجاع , ولا وجود للترتيب . يجب ان يكون $n \geq 1$.

مثال : بكم طريقة يمكن سحب (3) كرات من صندوق يحتوي على (7) كرات ؟

1. مع الارجاع ومراعات الترتيب .
2. مع الارجاع بدون ترتيب .
3. بدون ارجاع ومراعات الترتيب
4. بدون ارجاع وعدم مراعات الترتيب .

الحل:

a. عدد الطرق n^r :

b. $n^r = 7^3 = 7x7x7 = 343$

c. عدد الطرق C_r^{n+r-1}:

$$C_r^{n+r-1} = C_3^{7+3-1} = C_3^9 = \frac{9!}{3!(9-3)!} = 84$$

d. عدد الطرق P_r^n :

$$P_r^n = P_3^7 = \frac{7!}{3!(7-3)!} = 210$$

e. عدد الطرق C_r^n :

$$C_r^n = C_3^7 = \frac{7!}{3!(7-3)!} = 35$$

الفصل الحادي عشر

التوزيعات الاحتمالية Brobability Distribution

1. المتغير العشوائي Random Variables

المتغير العشوائي : نوع من التوزيعات الاحتمالية المهمة التي تدخل في كثير من المعالجات الاحصائية في مختلف انواع التخصصات العلمية , يتم هنا شرح مفهوم المتغير العشوائي بشيء من التفصيل :

تم تسمية المجموعة التي تتالف أو تحتوي على كل النتائج الممكنة لتجربة عشوائية (فضاء عينة) , كل نتيجة من هذه النتائج يمكن تمثيلها بنقطة أو ان كل نتيجة هي عبارة عن عنصر في فضاء العينة. من الممكن تمثيل كل نقطة او نتيجة او عنصر في فضاء العينة بعدد ما , وبذلك اصبح لدينا دالة معرفة على فضاء العينة. هذه الدالة يطلق عليها اسم (المتغير العشوائي Random Variables) , بمعنى ان نتائج التجربة العشوائية يصاحبها مقدار يسمى "المتغير العشوائي" عادة يرمز الى المتغير العشوائي بالرمز (y) .

نعني بالمتغير العشوائي هي ان قيمة المتغير تعتمد على نتائج التجربة فقط , بمعنى آخر ان نتائج التجربة العشوائية يرافقها دائماً مقدار متغير يسمى بالمتغير العشوائي , هذا المقدار ياخذ قيماً مختلفة حسب نتيجة التجربة العشوائية , أي ان هناك علاقة بين قيمة المتغير العشوائي (y) ونقاط فضاء العينة التي تصبح بعد اجراء التجربة والتي يرمز لها بالرمز (E) , وبذلك فان :

$$\boldsymbol{y = f\ (E)}$$

حيث ان:

y = المتغير العشوائي

f = دالة معرفة فضاء العينة

E = Event الحادث

اي اننا نهتم بقيمة المتغير العشوائي للتجربة بدلاً من جميع النتائج الممكنة للتجربة , وهو ما نطلق عليه (التوزيع الاحتمالي).

مثال (1): اذا تم رمي رمي قطعة نقود مرتين فان فضاء العينة سوف يكون لدينا اربعة احتمالات وكما يلي:

$$A = \boxed{\overset{\bullet}{HH} \quad \overset{\bullet}{HT} \quad \overset{\bullet}{TH} \quad \overset{\bullet}{TT}}$$

لنفترض ان المتغير العشوائي (y) يمثل عدد مرات ظهور الصورة , في هذه الحالة فان دالة أو قيمة المتغير العشوائي سوف تكون كما يلي:

فضاء العينة	(A)	H H	H T	T H	T T
قيمة المتغير العشوائي	(y)	2	1	1	0
التوزيع الاحتمالي		¼	¼	¼	¼

في هذه الحالة فان احتمال ظهور الصورة مرتين من هذه الاربع احتمالات يساوي (1) أو احتمال واحد ان تظهر الصورة مرتين. أي إن :

1. احتمال ظهور الصورة مرتين هو = $P\ (y = 2) = \frac{1}{4}$
2. احتمال ظهور الصورة مرة هو = $P\ (y = 1) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}$
3. احتمال عدم ظهور الصورة هو = $P\ (y = 0) = \frac{1}{4}$

إذا رمزنا لكل قيمة من قيم نتائج التجربة العشوائية بالرمز (y) وهو قيمة المتغير العشوائي وعليه فان:

$$y = 0\ ,1\ ,2$$

اي ان القيمة (صفر) مقترنة مع النقطة (T T) , وان القيمة (1) مقترنة مع النقطة (T H) و (H T) , وان القيمة (2) مقترنة مع الرقم (H H) , وهكذا حسب نتائج التجربة العشوائية. من هذا يتضح ان هماك علاقة بين قيمة المتغير العشوائي (y) ونقاط فضاء العينة , اي ان (y) هي دالة فضاء العينة.

مثال (2): تم القاء زهرتي طاولة مرة واحدة . بهذه الحالة فان التجربة العشوائية هي القاء زهرتين , وان نتائج التجربة هي النقاط التي تظهر على السطح العلوي للزهرتين, والمقدار الذي يرافق نتائج هذه التجربة هو الذي يكون مجموع النقاط التي تظهر على السطح العلوي للزهرتين , هذا المقدار ياخذ القيم حسب نقاط فضاء العينة وهي:

$$A = \{2\ ,3\ ,4\ ,5\ ,6\ ,7\ ,8\ ,9, 10\ ,11\ ,12\}$$

فضاء العينة كما موضح في الجدول التالي:

A =	1 , 1	2 , 1	3 , 1	4 , 1	5 , 1	6 , 1
	1 , 2	2 , 2	3 , 2	4 , 2	5 , 2	6 , 2
	1 , 3	2 , 3	3 , 3	4 , 3	5 , 3	6 , 3
	1 , 4	2 , 4	3 , 4	4 , 4	5 , 4	6 , 4
	1 , 5	2 , 5	3 , 5	4 , 5	5 , 5	6 , 5
	1 , 6	2 , 6	3 , 6	4 , 6	5 , 6	6 , 6

وبناء على ذلك فان مجموع النقاط التي تظهر على السطح العلوي للزهرتين هو متغير عشوائي , وهو متغير لانه ياخذ قيماً مختلفة حسب نتيجة التجربة العشوائية , وهو متغير عشوائي لانه يرافق نتائج تجربة عشوائية ولا يمكن معرفة قيمة المتغير العشوائي الا بعد اجراء التجربة .

مثال (3): اختيار طالب من بين طلاب الجامعة . التجربة العشوائية هي اختيار طالب . المطلوب هو (طول الطالب) , (دخل اسرة الطالب) , (عدد افراد اسرة الطالب) . نتائج التجربة هي عبارة عن متغير عشوائي لانه ياخذ قيماً مختلفة حسب نتيجة المتغير العشوائي .

ينقسم المتغير العشوائي الى قسمين :

1. المتغير العشوائي المنفصل Discrete Probability Distribution:

يقال ان المتغير العشوائي منفصل اذا اخذ قيما منفصلة عن بعضها البعض ,اي ياخذ قيماً صحيحة فقط, اي يوجد بينهما ثغرات . مثل عدد افراد الاسرة هو متغير منفصل لانه ياخذ القيم ..... ,5 , 4 , 3 , 2) etc) , وهذه القيم توجد بينهما ثغرات او فواصل , حيث لا توجد عائلة عدد افراد اسرتها يساوي **3½** فرد . كذلك مجموع النقاط التي تظهر على السطح العلوي عند القاء زهرة الطاولة هي متغير عشوائي منفصل لعدم وجود اجزاء من الارقام الصحيحة بين الارقام الطبيعية .

2. المتغير العشوائي المتصل Continuous Probability Distribution :

المتغير العشوائي المتصل هو المتغير الذي ياخذ عدد غير منته او اعداد غير منتهية , اي انه ياخذ جميع القيم التي تقع في نطاق تغيره, وهو الذي ياخذ جميع القيم الصحيحة والكسرية في مدى تغيره . مثال على ذلك طول الطالب , فهو متغير متصل لانه ياخذ قيما مستمرة في نطاق تغير طول الطالب . اذا كان طول الطالب يساوي مثلاً (160) سم , فانه ممكن ان يكون (160.5) سم , أو (160.3) سم , وتوجد قيم اخرى اقل للدقة تصل الى المليمترات أو اجزاء الملمتر , وهكذا .

مثال (1) : صندوق يحتوي على (4) كرات حمراء و (3) سوداء , سحبت منه كرتان , ما هو احتمال ان تكون الكرتان حمراء؟

الحل:

احتمال ان تكون الكرتان حمراء هو ان قيمة المتغير العشوائي (y) هو ظهور الكرة الحمراء . نرمز بهذه الحالة لنتائج التجربة بالرمز (y) , وعليه فان قيمة (y) سوف تكون حسب الجدول الاتي :

دالة فضاء العينة f(E)	القيمة (y)	P (y)
R R	2	$\frac{1}{4}$
R B	1	$\frac{1}{4}$
B R	1	$\frac{1}{4}$
B B	0	$\frac{1}{4}$

احتمال ان نحصل على الكرتان حمراء هو : $P\ (y = 2) = \frac{1}{4}$

احتمال ان نحصل على كرة واحدة حمراء هو : $P\ (y = 1) = \frac{2}{4}$

احتمال ان لا نحصل على كرة حمراء هو : $P\ (y = 0) = \frac{1}{4}$

مثال (2): تم رمي قطعة نقود ثلاث مرات , وان (y) يمثل عدد مرات ظهور الصورة (H) . أوجد احتمال عدد مرات ظهور الصورة في هذه التجربة العشوائية ؟

الحل :

يمكن توضيح احتمال ظهور الصورة من خلال الجدول التالي:

دالة فضاء العينة f(E)	قيمة المتغير العشوائي (y)	احتمال ظهور الصورة لكل حالة ممكنة		P(y)
H H H	3	1	احتمال ظهور الصورة مرة واحدة = 1 مرة	$P(y = 3) = \frac{1}{8}$
H T H	2	1	احتمال ظهور الصورة مرتين = 3 مرات	$P(y = 2) = \frac{3}{8}$
H H T	2	1		
T H H	2	1		
T T H	1	1	احتمال ظهور الصورة مرة واحدة = 3 مرات	$P(y = 1) = \frac{3}{8}$
H T T	1	1		
T H T	1	1		
T T T	0	1	احتمال عدم ظهور الصورة = 1 مرة	$P(y = 0) = \frac{1}{8}$

اذن مجموع الاحتمالات كافة يجب ان يساوي واحد:

$$\frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1$$

<u>مثال (3)</u>: تم القاء زهرتي طاولة مرة واحدة , أوجد التوزيع الاحتمالي لمجموع النقاط التي تظهر على السطح العلوي بعد اجراء التجربة ؟

الحل:

يتم رسم جدول يمثل نتائج التجربة وهو الذي يكون على اساس فضاء العينة وكما يلي:

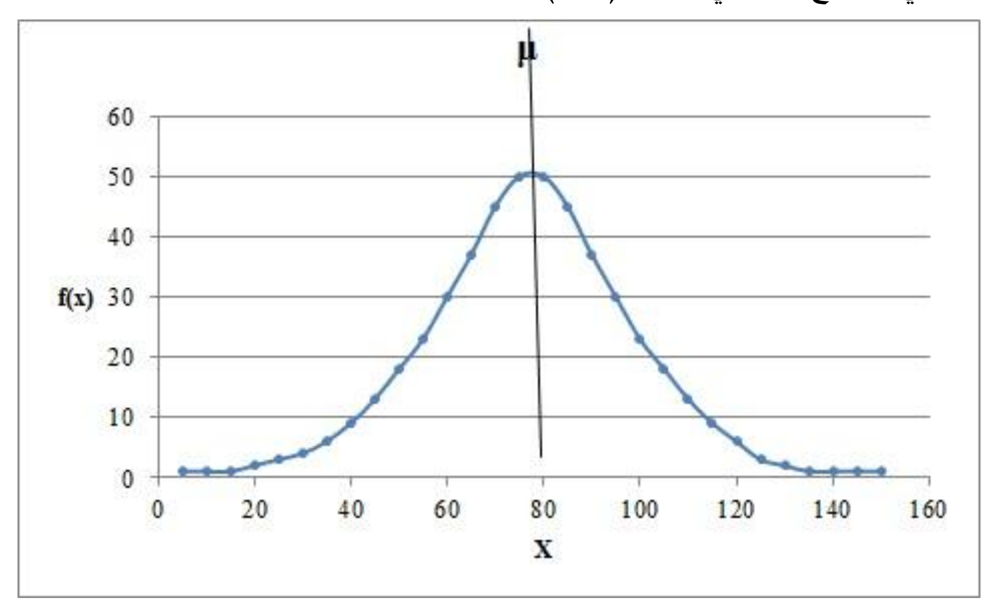
	1 , 1	2 , 1	3 , 1	4 , 1	5 , 1	6 , 1
ظهور الرقم 2 =1	1 , 2	2 , 2	3 , 2	4 , 2	5 , 2	6 , 2
ظهور الرقم 3 = 2	1 , 3	2 , 3	3 , 3	4 , 3	5 , 3	6 , 3
ظهور الرقم 4 = 3	1 , 4	2 , 4	3 , 4	4 , 4	5 , 4	6 , 4
ظهور الرقم 5 = 4	1 , 5	2 , 5	3 , 5	4 , 5	5 , 5	6 , 5
ظهور الرقم 6 = 5	1 , 6	2 , 6	3 , 6	4 , 6	5 , 6	6 , 6
ظهور الرقم 7 = 6	ظهور الرقم 8 = 5	ظهور الرقم 9 = 4	ظهور الرقم 10 = 3	ظهور الرقم 11 = 2	ظهور الرقم 12 = 1	

وبذلك يكون التوزيع الاحتمالي المطلوب هو:

y	2	3	4	5	6	7	8	9	10	11	12
f(y)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2. **التوزيع الطبيعي Normal Distribution:**

وهو من اهم التوزيعات المستمرة وتاخذ دالة كثافته الاحتمالية شكل منحني متماثل ذو قمة واحدة ويمتد طرفاه الى ما لا نهاية (∞), وقد وجد ان معظم التوزيعات الاحتمالية مثل : الاوزان, الاطوال, الاعمار , ونحو ذلك تاخذ شكلاً قريباً من منحني التوزيع الطبيعي, شكل (؟−1).



شكل (   ) منحني التوزيع الطبيعي

اذا كانت (X) متغير عشوائي متصل (مستمر), فان دالة كثافته الاحتمالية هي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad -\infty \leq X \leq \infty$$

بهذه الحالة يقال ان (X)يتبع توزيع طبيعي عادي متوسطه الحسابي هو (μ) وان انحرافه المعياري (σ) , ويكتب باختصار :

$$\boldsymbol{X \sim N\,(\mu\,,\sigma)}$$

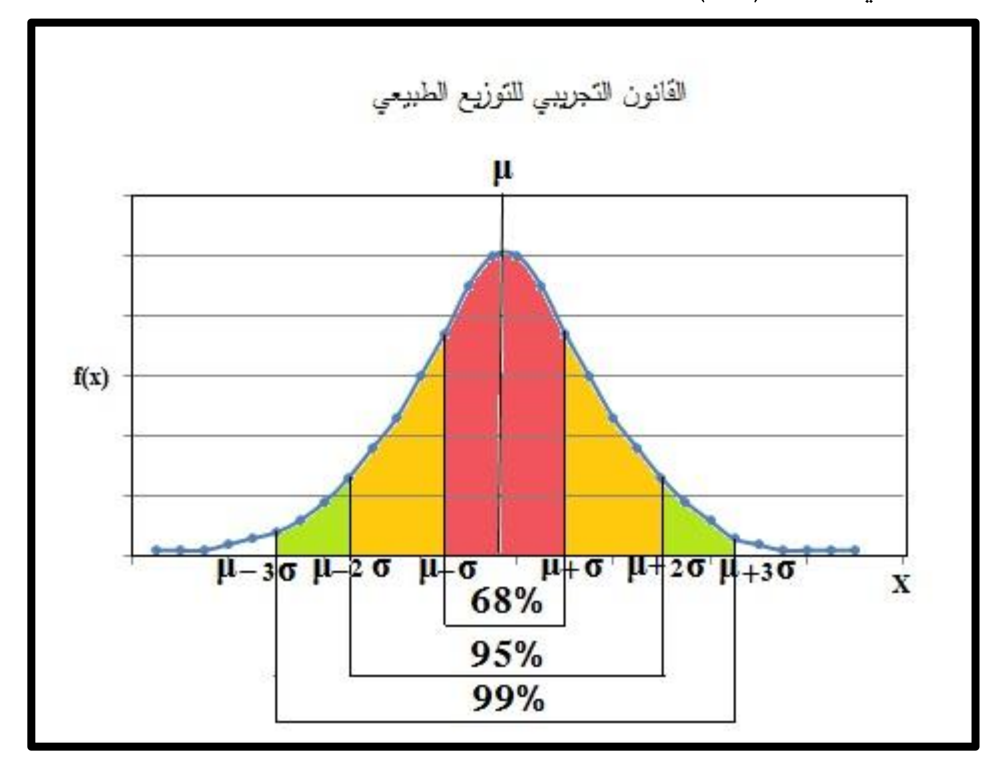
من خصائص التوزيع التوزيع الطبيعي انه متماثل حول متوسطه الحسابي, كما ان المساحة تحت منحني التوزيع تساوي (واحد), وحيث ان التوزيع متماثل فان المساحة تحت نصف المنحني تساوي (0.5).

القانون التجريبي للتوزيع الطبيعي هو:

1. ان 68% من البيانات تقع او تنحصر بين $(\mu \pm \sigma)$. اي يمين ويسار المتوسط الحسابي بمقدار واحد انحراف معياري.
2. ان 95% من البيانات تقع او تنحصر بين $(\mu \pm 2\sigma)$. اي يمين ويسار المتوسط الحسابي بمقدار اثنان انحراف معياري.

3. ان 99% من البيانات تقع او تنحصر بين $(\mu \pm 3\sigma)$. اي يمين ويسار المتوسط الحسابي بمقدار ثلاثة انحراف معياري.

هذه النسب كما مؤشرة في الشكل ( ).



شكل ( ) القانون التجريبي للتوزيع الطبيعي

<u>التوزيع الطبيعي القياسي:</u>

يمكن ان نقول ان المتغير العشوائي المتصل الذي نرمز له بالرمز (Z) , يتبع توزيع قياسي طبيعي متوسطه الحسابي هو $(\mu = 0)$, وان انحرافه المعياري هو $(\sigma = 1)$, ويكتب باختصار كما في العلاقة التالية:

$$Z \sim N(0,1)$$

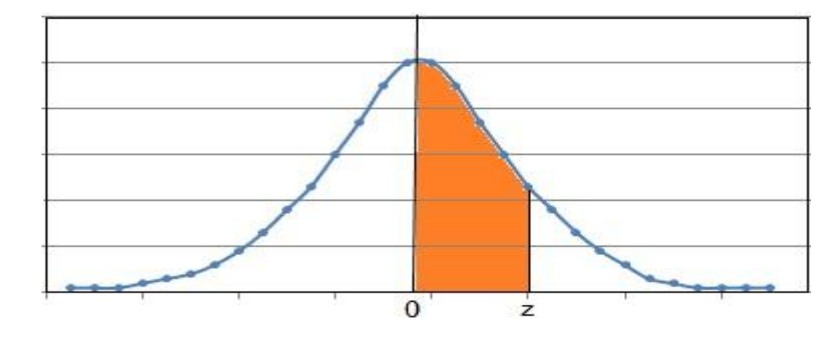
اذا كان المتغير العشوائي المتصل (Z) يعرف بموجب العلاقة التالية : $Z = \frac{X-\mu}{\sigma}$, ومن خلال التوزيع الطبيعي القياسي يمكن حساب قيمة اي احتمال لمتغير عشوائي يتبع توزيع طبيعي عادي الذي يمكن توضيحه بالعلاقة التالية: $P(0 < X < a) = P(-a < X < 0)$.

لقد خصصت جداول قياسية توضح قيمة الاحتمال تحت منحني التوزيع الطبيعي القياسي تسمى جدول (Z), الموضح كما يلي:

جدول التوزيع الطبيعي القياسي

$Z \sim N(0,1)$

المساحة المظللة تمثل $P(0 < Z < z)$



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.04	0.08	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.5331	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.479	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499
3.1	0.499	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997

شروط استخدام جدول (Z):

1. اذا كان السؤال في التوزيع الطبيعي العادي , بمعنى ان$\sigma \neq 0$, $\mu \neq 0$, فاننا بهذه الحالة يجب ان نحول التوزيع الطبيعي العادي الى قياسي باستخدام تعرف (Z).
2. ان جميع قيم الاحتمالات في الجدول محسوبة للنصف الموجب , وما يقال على النصف الموجب ينطبق على النصف السالب بسبب تماثل التوزيع.
3. الاحتمالات محسوبة جميعاً على الصورة $P(0 < Z < z)$.

مثال(1):

اذا كان دخل (1000) اسرة في مدينة معينة يتبع التوزيع الطبيعي الذي متوسطه الحسابي (μ) يساوي (1800) دينار, وانحرافه المعياري (σ) يساوي (300) دينار . أوجد:

1. احتمال الحصول على دخل اسرة اكبر من 2400 دينار.
2. احتمال الحصول على دخل اسرة اكبر من 1500 دينار.
3. احتمال الحصول على دخل اسرة اقل من 2550 دينار.
4. احتمال الحصول على دخل اسرة اقل من 1200 دينار.
5. احتمال الحصول على دخل اسرة ينحصر بين 1650 و 2250 دينار.
6. احتمال الحصول على دخل اسرة اكبر من 2400 و 2550 دينار
7. اوجد عدد الاسر التي يزيد دخلها عن 1500 دينار

الحـــــــل:

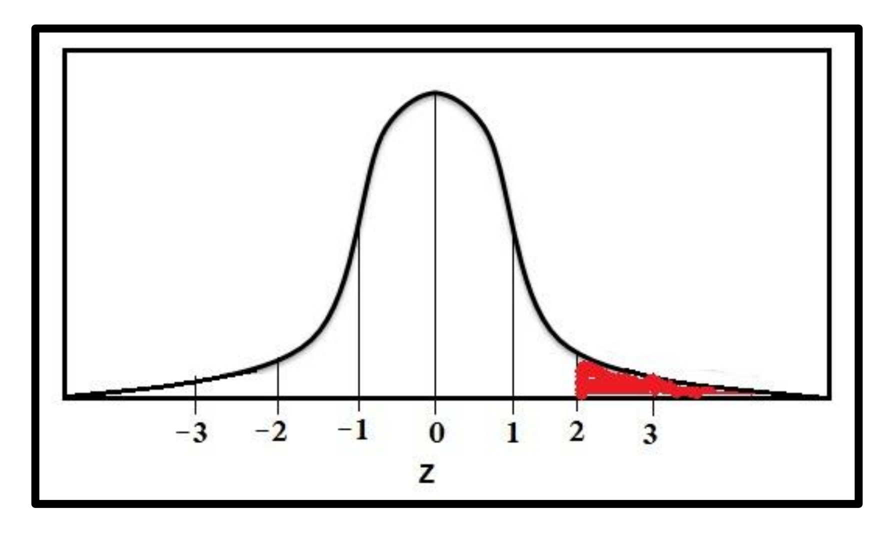
بما ان المتغير العشوائي (X) يتبع توزيع طبيعي عادي حسب منطوق السؤال, وان متوسطه الحسابي (μ) يساوي 1800 دينار, وان الانحراف المعياري لهذا المتغير (σ) يساوي 300 دينار, وعليه فان:

1. احتمال ان يكون دخل اسرة اكبر من 2400 دينار:

$$P(X \geq 2400) = P\left[\left(\frac{X-\mu}{\sigma}\right) \geq \frac{2400-\mu}{\sigma}\right] = P\left(Z \geq \frac{2400-\mu}{\sigma}\right) = P\left(Z \geq \frac{2400-1800}{300}\right) = P(Z \geq 2)$$

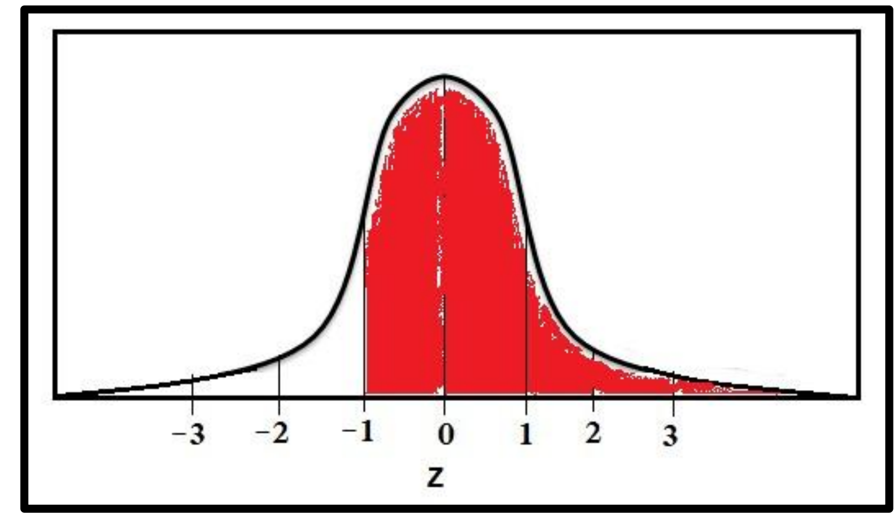
$$P(X \geq 2400) = P(Z \geq 2) = 0.5 - \underbrace{P(0 \leq Z \leq 2)}_{\text{من الجدول Z}} = 0.5 - 0.4772 = 0.0228$$

كما هو موضح في الشكل التالي:



2. احتمال ان يكون دخل اسرة اكبر من 1500 دينار:

$$P(X \geq 1500) = P\left[\left(\frac{1500-\mu}{\sigma}\right)\right] = P\left(Z \geq \frac{1500-1800}{300}\right) = P(Z \geq -1)$$

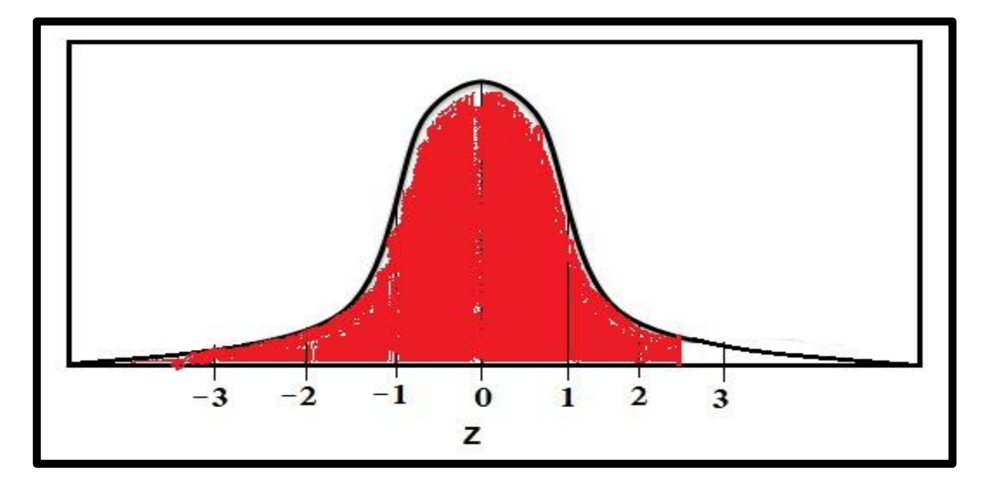


$$P(X \geq 1500) = P(Z \geq -1) = 0.5 + P(-1 \leq Z \leq 0) = 0.5 + \underbrace{P(0 \leq Z \leq 1)}_{\text{من الجدول}} =$$
$$0.5 + 0.3413 = 0.8413$$

3. احتمال ان يكون دخل اسرة اقل من 2550 دينار:

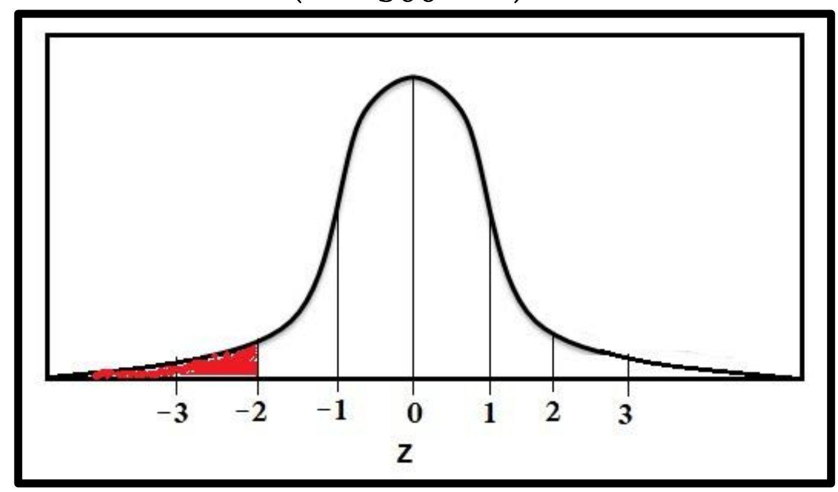
$$P(X \leq 2550) = P\left(Z \leq \frac{2550-1800}{300}\right) = P(Z \leq 2.5)$$

$$\mathrm{P}(X \leq 2550) = P(Z \leq 2.5) = 0.5 + P(0 \leq Z \leq 2.5) = 0.5 + 0.4938 = 0.9938$$



4. احتمال ان يكون دخل اسرة اقل من 1200 دينار:

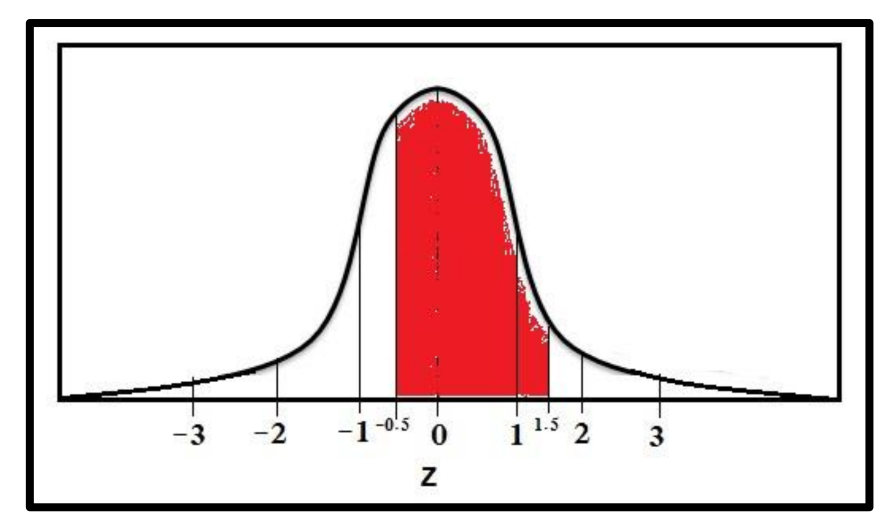
$$P(X \leq 1200) = P\left(\frac{1200 - 1800}{300}\right) = P(Z \leq -2)$$



$$P(X \leq 1200) = P(Z \leq -2) = 0.5 - P(0 \leq Z \leq 2) = 0.5 + 0.4772 = 0.0228$$

5. احتمال ان يكون دخل اسرة ينحصر بين 1650 و 2250 دينار:

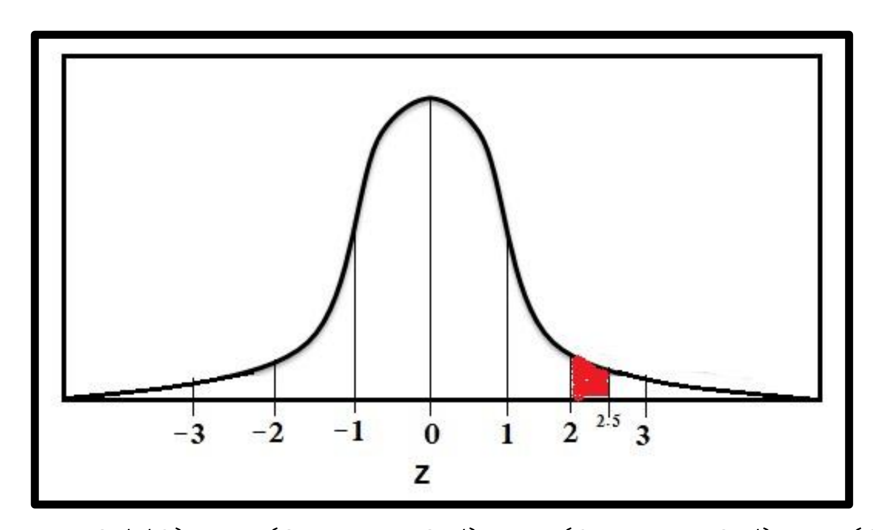
$$P(1650 \leq X \leq 2250) = P\left(\frac{1650 - 1800}{300} \leq Z \leq \frac{2250 - 1800}{300}\right)$$
$$= P(-0.5 \leq Z \leq 1.5)$$



$$P(1650 \le X \le 2250) = P(-0.5 \le Z \le 1.5) = P(0 \le Z \le 0.5) + P(0 \le Z \le 1.5)$$
$$= 0.1915 + 0.4332 = 0.6247$$

6. احتمال ان يكون دخل اسرة ينحصر بين 2400 و 2550 دينار:

$$P(2400 \le X \le 2550) = P\left(\frac{2400 - 1800}{300} \le Z \le \frac{2550 - 1800}{300}\right)$$
$$= P(2 \le Z \le 2.5)$$



$$P(2400 \le X \le 2550) = P(2 \le Z \le 2.5) = P(0 \le Z \le 2.5) - P(0 \le Z \le 2)$$
$$= 0.4938 - 0.4772 = 0.0166$$

7. عدد الاسر التي يزيد دخلها عن 1500 دينار:

ان احتمال عدد الاسر التي يزيد دخلها عن (1500) دينار هي :

(الاحتمال الناتج من الفقرة 2 مضروباً بعدد الاسر الكلي وهو (1000) اسرة) = 0.8413 X 1000

=841 اسرة

الفصل الثاني عشر

نظرية ذي الحدين Binomial Distribution

وتسمى كذلك توزيع ذي الحدين , ويسمى ايضاً توزيع برنولي Bernoulli Distribution , الذي اكتشفه العالم برنولي في القرن السابع عشر, ويسمى كذلك بالتوزيع الحداني .

في كثير من التجارب يعمد الباحث على تكرار اعادة اجراء التجربة في عدد كبير من المرات وذلك لرصد وتسجيل نجاح او فشل ظاهرة معينة , وتسمى مثل هذه التجارب , (تجارب ذي الحدين) , وذلك لانها تركز فقط على نتيجتين فقط هما (حالة النجاح و حالة الفشل) , وتحديد عدد مرات ظهور النتيجة المطلوبة (النجاح) من العدد الكلي لمرات اجراء التجربة.

توزيع ذي الحدين من التوزيعات الاحتمالية المتقطعة المهمة والذي يكون فيه المتغير العشوائي يقبل (1) توزيعاً له اي ان : (عدد حالات النجاح + عدد حالات الفشل =1) . يمكن توضيح هذا التوزيع الاحتمالي باستخدام بعض الامثلة :

نفترض ان لدينا تجربة ما , بحيث ان جميع نتائج هذه التجربة يمكن تصنيفها الى صنفين أو حالتين وهي اما حالة نجاح أو حالة فشل , أي ان ظهور الحادث (A) مثلاً , نطلق عليه حالة النجاح , وعدم ظهور الحادث (A) نطلق عليه حالة الفشل . نفترض ان هذه التجربة تتكرر عدد من المرات ولتكن (n) من المرات. نفترض ان المتغير العشوائي (y) يمثل عدد حالات النجاح (اي عدد مرات ظهور الحادث (A)) , التي تظهر في عدد مرات تكرار التجربة (n) .

هذا النوع من المتغير العشوائي يسمى (متغير ذي الحدين). وهو متقطع لانه ياخذ قيماً عددية من الصفر الى (n) . إن تكرار التجربة في هذه الحالة يكون تكرار لاصل التجربة في كل مرة , اي ان التجارب المتكررة تكون مستقلة . أي ان الحادث الذي يظهر في التجربة الاولى لا يغير من الاختيار الذي يليه في التجربة الثانية.

مثال (1) : في حالة رمي قطعة نقود ثلاث مرات (اي ان التجربة يتم تكرارها ثلاث مرات) وعليه فان: $(n = 3)$. نفترض ان حالات النجاح هو الحصول على صورة (H) , وبذلك فان المتغير العشوائي (y) يمثل عدد مرات ظهور الصورة أو التي نحصل عليها من هذه الرميات الثلاثة . التوزيع الاحتمالي لمتغير ذي الحدين في هذه الحالة يكون كما يلي :

فضاء العينة	(y) قيمة المتغير العشوائي (عدد مرات ظهور الحادث)	احتمال كل حالة ممكنة	P(E) توزيع ذي الدين (مجموع الحالات الممكنة)
H H H	3	$\frac{1}{8}$	$\frac{1}{8}$
T T T	0	$\frac{1}{8}$	$\frac{1}{8}$
H T T	1	$\frac{1}{8}$	$\frac{3}{8}$
T T H	1	$\frac{1}{8}$	
T H T	1	$\frac{1}{8}$	
H H T	2	$\frac{1}{8}$	$\frac{3}{8}$
T H H	2	$\frac{1}{8}$	
H T H	2	$\frac{1}{8}$	
المجموع		1.0	1.0

1) في هذا المثال , فان فضاء العينة مكون من ثمانية احتمالات .
2) كل احتمال يحصل مرة واحدة من ثمانية احتمالات.
3) عدد حالات النجاح هو ظهور الصورة , عندئذ يتم جمع عدد احتمالات ظهور الصورة ضمن فضاء العينة.
4) عدد حالات النجاح + عدد حالات الفشل = 1

طريقة أو خطوات ايجاد توزيع ذي الحدين:

1. ايجاد فضاء العينة (وهي جميع النتائج الممكنة).
2. ايجاد احتمالية حصول كل نتيجة من النتائج الممكنة او المحتملة .
3. تعيين قيمة المتغير العشوائي .
4. جمع كافة احتمالات النجاح , هذه الاحتمالات تعطي أو تمثل التوزيع الاحتمالي لتوزيع ذي الحدين .

القانون العام لتوزيع ذي الحدين :

1. نرمز الى احتمال ان يحصل الحادث (احتمال النجاح) بالرمز (p) .
2. نرمز الى احتمال عدم حصول الحادث (احتمال الفشل) بالرمز (q) .
3. بهذه الحالة يكون : $\boldsymbol{p+q=1}$. اي ان :$\boldsymbol{q=1-p}$ وكذلك : $\boldsymbol{p=1-q}$.
4. بهذه الحالة يكون الاحتمال ان تحصل حالات النجاح في (X) مرة من عدد مرات اجراء التجربة في (n) من عدد مرات تكرار التجربة .
5. توزيع ذي الحدين يعطى بالعلاقة التالية :

$$\boldsymbol{p}_{(y=x)=\binom{n}{x}}\ x\ p^{X}\ x\ q^{n-X} = \frac{n!}{X!(n-X)!}\ x\, p^{X}\ x\ q^{n-X}$$

حيث ان :

X = عدد حالات النجاح

n = عدد مرات اجراء التجربة

p = احتمال النجاح في كل تجربة

q = احتمال الفشل في كل تجربة

p(y) = المتغير العشوائي بعدد مرات النجاح , y = 0 , 1 , 2 , 3 , 4 ….. etc.

نعود الى المثال السابق عند رمي قطعة نقود ثلاث مرات ونطبق قانون ذي الحدين :

n = 3

p = ½ , q = ½

X = 0 ,1 , 2 , 3 عدد مرات ظهور الصورة

نطبق قانون ذي الحدين:

$$p(y=x)=\binom{n}{x}\ x\ p^{X}\ x\ q^{n-X}$$

1. $\mathbf{p(y=0)=\binom{3}{0}\ x\ (½)^{0}\ x\ (½)^{3-0}=\frac{1}{8}}$ عدم ظهور الصورة
2. $\mathbf{p(y=1)=\binom{3}{1}\ x\ (½)^{1}\ x\ (½)^{3-1}=\frac{3}{8}}$ ظهور الصورة مرة واحدة
3. $\mathbf{p(y=2)=\binom{3}{2}\ x\ (½)^{2}\ x\ (½)^{3-2}=\frac{3}{8}}$ ظهور الصورة مرتين
4. $\mathbf{p(y=3)=\binom{3}{3}\ x\ (½)^{3}\ x\ (½)^{3-3}=\frac{1}{8}}$ ظهور الصورة ثلاث مرات

مثال (1) : ما هي احتمالات الحصول على وجه (H) مرتين فقط عند رمي قطعة نقود ستة مرات ؟

الحل :

$n = 6$

$p = ½$, $q = ½$

$y = 2$

نطبق قانون ذي الحدين:

$$p(y=x)=\binom{n}{x}\ x\ p^{X}\ x\ q^{n-X}$$

1. $\mathbf{p(y=2)=\binom{6}{2}\ x\ (½)^{2}\ x\ (½)^{6-2}=\frac{6!}{2!(6-2)!}\ x\ \left(\frac{1}{2}\right)^{2}\ x\ \left(\frac{1}{2}\right)^{4}=\frac{15}{64}}$

مثال (2) : في النية حفر خمسة آبار لاستكشاف النفط في احدى المناطق , حيث ان نسبة النجاح المتوقعة هي (10%) في الحصول على آبار ناجحة. احسب احتمالية :

A. فشل برنامج الحفر الاستكشافي كلياً .

B. الحصول على بئر واحد فقط يعطي مؤشر على وجود النفط .

الحل :

.One العدد الكلي للآبار $n = 5$

عدد حالات النجاح $X = 0$ لعدم وجود بئر يعطي مؤشر لتواجد النفط

احتمال النجاح $p = 10\% = 0.01$

احتمال الفشل $q = 1 - 10\% = 90\% = 0.90$

نطبق القانون $p(y = x) = \binom{n}{x}\ x\ p^{X}\ x\ q^{n-X}$

$$p(y = 0) = \binom{5}{0}\ x\ (0.10)^{0}\ x\ (0.90)^{5-0} = \frac{5!}{0!\,(5-0)!}\ \text{x}\ (0.10)^{0}\ \text{x}\ (0.9)^{5}$$
$$=$$
$$1\ x\ 1\ x\ 0.59 = 59\%$$

احتمال فشل برنامج الحفر الاستكشافي كلياً

.Two العدد الكلي للآبار $n = 5$

عدد حالات النجاح $X = 1$ بئر واحد يعطي مؤشر لتواجد النفط

احتمال النجاح $p = 10\% = 0.01$

احتمال الفشل $q = 1 - 10\% = 90\% = 0.90$

نطبق القانون $p(y = x) = \binom{n}{x}\ x\ p^{X}\ x\ q^{n-X}$

$$p(y = 0) = \binom{5}{1}\ x\ (0.10)^{1}\ x\ (0.90)^{5-1} = \frac{5!}{1!\,(5-1)!}\ \text{x}\ (0.10)^{1}\ \text{x}\ (0.9)^{1}$$
$$=$$
$$5\ x\ 0.1\ x\ 0.656 = 38\%$$

نسبة النجاح في الحصول على بئر واحد ناجح فقط

مثال (2) : ما هي احتمالية الحصول على وجه (H) على الاقل اربعة مرات عند رمي قطعة نقود ستة مرات؟

الحل :

$$y = 4\,, \quad y = 5\,, \quad y = 6\,, \quad n = 6$$

على الاقل , تعني ان الاحتمالية هي الحصول على وجه اربع مرات وخمس مرات وست مرات

$$p = \frac{1}{2} \quad , \quad q = \frac{1}{2}$$

نطبق القانون $p(y = x) = \binom{n}{x}\ x\ p^{X}\ x\ q^{n-X}$

$$p(y = 4 + 5 + 6)$$

$$= \binom{6}{4}\ x\ \left(\frac{1}{2}\right)^{4}\ x\ \left(\frac{1}{2}\right)^{6-4} + \binom{6}{5}\ x\ \left(\frac{1}{2}\right)^{5}\ x\ \left(\frac{1}{2}\right)^{6-5}$$

$$+ \binom{6}{6}\ x\ \left(\frac{1}{2}\right)^{6}\ x\ \left(\frac{1}{2}\right)^{6-6} = \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{11}{32}$$

مثال (3) : اذا كان احتمال اصابة لاعب كرة السلة (A) الهدف في رمية حرة هو $\frac{3}{4}$, فما هو احتمال اصابة الهدف مرتين من (4) رميات حرة .

الحل :

$$p = \frac{3}{4} \quad , \quad q = 1 - q = 1 - \frac{3}{4} = \frac{1}{4} \quad ,$$

$$n = 4 \quad , \quad X = 2$$

$$P = (y = 2) = \binom{4}{2} x\ \frac{3^2}{4}\ x\ \frac{1}{4}^{4-2} = \frac{4!}{2!\,(4-2)!}\ x\ \frac{3^2}{4}\ x\ \frac{1^2}{4} = 6\ x\ \frac{9}{16}\ x\ \frac{1}{16}$$

$$= \frac{54}{94}$$

مثال (4) : وجد في احد المصانع ان نسبة العلب التالفة من معجون الطماطة التي ينتجها المصنع هي (5%) , فاذا اخذت عينة مؤلفة من (10) علب , احسب احتمال :

a. ان تكون العينة كلها تالفة .

b. ان تكون العينة كلها جيدة .

c. ان تكون بالعينة ثلاث علب تالفة فقط .

الحل :

a. $p = 5\% = 0.05 \quad , \quad q = 1 - 0.05 = 0.95 \quad , \quad n = 10 \quad , \quad X = 10$

$$P = (y = 10) = \binom{10}{10} x\ (0.05)^{10}\ x\ (0.95)^{10-10}$$

b. $p = 95\% = 0.95 \quad , \quad q = 1 - 0.95 = 0.05 \quad , \quad n = 10 \quad , \quad X = 10$

$$P = (y = 10) = \binom{10}{10} x\ (0.95)^{10}\ x\ (0.05)^{10-10}$$

c. $p = 5\% = 0.05 \quad , \quad q = 1 - 0.05 = 0.95 \quad , \quad n = 10 \quad , \quad X = 10$

$$P = (y = 3) = \binom{10}{3} x\ (0.05)^{3}\ x\ (0.95)^{10-3}$$

مثال (5) : وجد عند مفترق طريقين ان $\left(\frac{2}{3}\right)$ من السيارات تتجه الى اليمين , وان $\left(\frac{1}{3}\right)$ منها تتجه الى اليسار . فاذا وصلت (4) سيارات الى المفترق , فما هو احتمال ان تتجه (3) منها الى اليسار ؟

الحل :

احتمال النجاح هو اتجاه السيارات الى اليسار = $\boldsymbol{p = \frac{1}{3}}$,

احتمال الفشل هو اتجاه السيارات الى اليمين = $\boldsymbol{q = 1 - \frac{1}{3} = \frac{2}{3}}$

عدد حالات النجاح = $\boldsymbol{X = 3}$, عدد مرات اجراء التجربة = $\boldsymbol{n = 4}$

$$p(y = 3) = \binom{4}{3}\ x\ \left(\frac{1}{3}\right)^{3}\ x\ \left(\frac{2}{3}\right)^{4-3} = \frac{4!}{3!\,(4-3)!}\ x\ \left(\frac{1}{3}\right)^{3}\ x\ \left(\frac{2}{3}\right)^{4-3}$$
$$= 4x0.037x0.666 = 0.098$$

مثال (6) : في احدى تجارب مندل الوراثية , وجد بان احتمال الحصول على نبات طويل يساوي $\left(\frac{3}{4}\right)$, وعلى نبات قصير يساوي $\left(\frac{1}{4}\right)$ في الجيل الثاني , فاذا فحصت عينة مؤلفة من (4) نباتات , فما هو احتمال ان تكون كلها طويلة ؟

الحل :

احتمال النجاح هو = $\boldsymbol{p = \frac{3}{4}}$,

احتمال الفشل هو = $\boldsymbol{q = 1 - \frac{3}{4} = \frac{1}{4}}$

عدد حالات النجاح = $X = 4$, عدد مرات اجراء التجربة = $n = 4$

$$p(y=4) = \binom{4}{4} \, x \left(\frac{3}{4}\right)^4 \, x \left(\frac{1}{4}\right)^{4-4} = \frac{4!}{4!\,(4-4)!} \, x \left(\frac{3}{4}\right)^4 \, x \left(\frac{1}{4}\right)^0 = \frac{81}{256}$$

مثال (7) : وجد ان من بين كل عشرة نماذج طين , هنالك نموذج واحد فقط يحتوي على تراكيز للحديد , باستخدام توزيع ذي الحدين , اوجد احتمال ان يكون من بين كل (20) نموذج طين هناك (3) نماذج تحتوي على تراكيز للحديد ؟

الحل :

$$p = 3 \;,\; q = 1 - 3 = -2 \,, \quad n = 20 \,, \quad X = 3$$

$$p(y=3) = \binom{20}{3} \, x \, (3)^3 \, x \, (-2)^{20-3}$$

خواص توزيع ذي الحدين :

1. الوسط الحسابي : وهو عبارة عن معدل عدد النجاحات المتوقعة التي يمكن الحصول عليها في (n) من الحالات او التجارب . ويرمز له بالرمز (μ) أو $E(y)$. والصيغة الاحصائية لاحتساب الوسط الحسابي لمتغير ذي الحدين هي :

a) $\mu = E(y) = \sum_{y=0}^{n} y \, [p(y)]$

b) $\mu = n \, x \, p$

حيث ان :

المتغير العشوائي وهو عدد مرات ظهور الحادث او عدد حالات النجاح = y

عدد مرات اجراء التجربة = n

مجموع الاحتمالات في فضاء العينة = $p(y)$

احتمال ظهور الحادث في كل مرة اجراء التجربة (حالة النجاح) = p

مثال : أوجد الوسط الحسابي لتجربة رمي قطعة نقود (4) مرات , والحادث هو ظهور الصورة في عدد مرات اجراء التجربة .

الحل :

مجموع الاحتمالات في فضاء العينة = $16 = 2 \times 2 \times 2 \times 2$ وان احتمال ظهور الصورة موضحة في الجدول التالي :

توزيع ذي الحدين	احتمال كل حالة ممكنة	قيمة المتغير العشوائي (y)	فضاء العينة
$\frac{1}{16}$ احتمال عدم ظهور الصورة	$\frac{1}{16}$	0	T T T T
$\frac{4}{16}$ احتمال ظهور الصورة مرة واحدة	$\frac{1}{16}$	1	H T T T
	$\frac{1}{16}$	1	T H T T
	$\frac{1}{16}$	1	T T H T
	$\frac{1}{16}$	1	T T T H
$\frac{6}{16}$ احتمال ظهور الصورة مرتين	$\frac{1}{16}$	2	H H T T
	$\frac{1}{16}$	2	H T H T
	$\frac{1}{16}$	2	H T T H
	$\frac{1}{16}$	2	T H T H
	$\frac{1}{16}$	2	T H H T
	$\frac{1}{16}$	2	T T H H
$\frac{4}{16}$ احتمال ظهور الصورة ثلاث مرات	$\frac{1}{16}$	3	T H H H
	$\frac{1}{16}$	3	H T H H
	$\frac{1}{16}$	3	H H T H
	$\frac{1}{16}$	3	H H H T
$\frac{1}{16}$ احتمال ظهور الصورة اربع مرات	$\frac{1}{16}$	4	H H H H

نطبق القانون الاول :

$$\boldsymbol{\mu = E(y) = \sum_{y=0}^{n} y\,[p(y)] = 0x\left(\frac{1}{16}\right) + 1x\left(\frac{4}{16}\right) + 2x\left(\frac{6}{16}\right) + 3x\left(\frac{4}{16}\right) + 4x\left(\frac{1}{16}\right)}$$

$$= \frac{0}{16} + \frac{4}{16} + \frac{12}{16} + \frac{12}{16} + \frac{4}{16} = 2$$

بما ان هذه الطريقة طويلة خاصة اذا كان عدد النقاط في فضاء العينة كبير جداً , بهذه الحالة يفضل تطبيق القانون الثاني :

$$\mu = n\ x\ p = 4\ x\ \frac{1}{2} = 2$$

2. التباين : القانون الخاص بايجاد التباين هو :

$$\sigma^2 = n\ x\ p\ x\ q$$

3. الانحراف المعياري :

$$\sigma^2 = \sqrt{n\ x\ p\ x\ q}$$

4. معامل الالتواء :

$$\alpha = \frac{q - p}{\sqrt{n\ x\ p\ x\ q}}$$

5. معامل التفرطح :

$$\beta = 3 + \frac{1 + (6\ x\ p\ x\ q)}{(n\ x\ p\ x\ q)}$$

الفصل الثالث عشر

الاحتمال الشرطي Conditional Probability

هناك كثير من الحوادث التي يشترط وقوعها بوقوع حوادث سابقة لها (تسبقها في الحدوث).

مثال : سفر الطالب للدراسة في الخارج مرتبط أو مشروط بنجاحه بامتحان القبول .

الحادث الاول : سفر الطالب الى الخارج

الحادث الثاني : نجاح الطالب في الامتحان

نلاحظ هنا ان الحادث الاول يقع بعد وقوع الحادث الثاني .فاذا رمزنا الى الحادث الاول (E_1) , والحادث الثاني (E_2) , وان الحادث الاول E_1 يقع بشرط وقوع الحادث الثاني E_2 وهذا يعبر عنه رياضياً $\boldsymbol{E1/E2}$ حيث تقرأ كما في احد الحالات التالية :

a. يقع الحادث الاول (E_1) بشرط ان الحادث الثاني (E_2) قد وقع .

b. (E_1) الحادث الاول يقع علما بان (E_2) قد وقع.

c. (E_1) يقع اذا كان (E_2) قد وقع.

d. (E_1) يقع على فرض ان (E_2) قد وقع.

وسنتعرف على كيفية ايجاد احتمال الحادث المشروط بوقوع حادث قبله .

اذا كان لدينا حادثين هما (E_1) و (E_2) , فان احتمال وقوع الحادث (E_2) (الحادث الثاني) مع ان الحادث الاول (E_1) قد وقع , نرمز له بهذه الحالة بالرمز : $\boldsymbol{p_{r=(E_2/E_1)}}$, ويسمى هذا بالاحتمال الشرطي , ويعني ذلك ان احتمال وقوع الحادث هو (E_2) مع ان الحادث (E_1) قد وقع مسبقاً أو تحقق حصوله .

<u>يقسم الاحتمال الشرطي الى ثلاثة اقسام :</u>

1) <u>الحوادث المستقلة Independent Events :</u>

وهذا يعني ان حصول الحادث (E_1) أو عدم حصول الحادث (E_1) , اذا كان ذلك لا يؤثر على وقوع الحادث (E_2) عندئذ سوف يكون :

$$\boldsymbol{p_{r=(E_2/E_1)} = p_{r=(E_2)}}$$

2) الحوادث غير المستقلة Dependent Events :

وهي الحوادث التي يشترط حصول الحادث (E_1) بوقوع الحادث (E_2) أو تحقق حصوله ويرمز له :

$$p_r(E_1\, E_2) = p_r\,(E_1)\ x\ p_r\,(E_2)$$

3) الحوادث المركبة Complex Events :

في حالة وجود حادثين (E_1 و E_2) وكلاهما حاصل الحدوث , عندئذ تسمى هذه الحوادث بالحوادث المركبة , وتعرف بالصيغة التالية :

$$p_r\,(E_1E_2) = p_r(E_1) \quad x \quad p_r\,(E_2/E_1)$$

احتمال حصول الحادث الاول — احتمال حصول الحادث الثاني

اما في حالة وجود عدة حوادث مستقلة :

$$p_r\,(E_1\, E_2\, E_3) = p_r\,(E_1)\ x\ \ p_r\,(E_2)\ \ \mathrm{x}\ \ p_r\,(E_3)\ \mathrm{x} \ldots ....$$

مثال (1) : في تجربة سحب كرتين من صندوق فيه (5) كرات بيضاء و (7) كرات سوداء و (3) كرات حمراء , اذا كان السحب على التوالي وبدون ارجاع , أوجد :

a) احتمال ان تكون الكرة الاولى بيضاء .

b) احتمال ان تكون الكرة الثانية بيضاء , اذا كانت الكرة الاولى بيضاء .

c) احتمال ان تكون الكرة الثانية سوداء , اذا كانت الكرة الاولى حمراء .

d) احتمال ان تكون الكرة الاولى بيضاء والثانية بيضاء ايضاً .

الحل :

في هذا المثال يتم السحب على التوالي بمعنى ان الترتيب مهم وبالتالي تصبح التجربة مكونة من خطوتين تتميزين بان حدوث السحبة الثانية مشروطة بحدوث وتحقق السحبة الاولى قبلها (احتمال مشروط) , وبالتالي سيكون لزاماً علينا دراسة احتمال السحبة الثانية بعد ان تعطى معلومات عن نتائج السحبة الاولى واعطاء مجريات نتائج وقوع السحبة الثانية , لان الترتيب مهم .

a) احتمال ان تكون الكرة الاولى بيضاء = $\frac{\text{عدد الكرات البيضاء}}{\text{عدد الكرات الكلي}} = \frac{\text{الحالات المؤاتية}}{\text{عدد الكرات الكلي}} = \frac{5}{15} = \frac{1}{3}$

b) احتمال ان تكون الكرة الثانية بيضاء (E_2) اذا كانت الاولى بيضاء (E_1) .

$$p_r = (E_2/E_1) = \frac{\text{عدد الكرات البيضاء بعد ان يكون ناتج السحبة الاولى كرة بيضاء}}{\text{عدد الكرات الكلي بعد نقصان كرة بيضاء}} = \frac{4}{14}$$

c) احتمال الكرة الثانية سوداء اذا كانت الكرة الاولى حمراء

$$p_r = (E_2/E_1) \ \frac{\text{عدد الكرات السوداء بعد ان تكون ناتج السحبة الاولى كرة حمراء}}{\text{عدد الكرات الكلي بعد نقصان كرة حمراء}} = \frac{7}{14} = \frac{1}{2}$$

d) احتمال ان تكون الكرة الاولى بيضاء والثانية بيضاء , وهي من الحوادث المركبة التي نطبق عليها القانون التالي :

$$p_r\,(E_1\,E_2) = p_r\,(E_1)\,x\,p_r\,(E_2/E_1) = \frac{5}{15}\,x\,\frac{4}{14} = \frac{4}{42}$$

مثال (2) : عينة مكونة من (20) طالب و (30) معلم , شاركوا في الاجابة عن اهمية استهلاك الطاقة في المنازل , فكانت اجابتهم كما ياتي :

الاجابة	نعم	لا	غير متاكد	المجموع
طلاب	14	4	2	20
معلمون	24	3	3	30

فاذا اختير احد افراد العينة عشوائياً , فما هو احتمال ان معلما علما بان اجابته كانت نعم .

الحل :

هنا الاختيار يجب ان يكون مشروط بان تكون اجابته (نعم) , اي ان الحادث الاول هو اجابته (نعم) = E_1 , والحادث الثاني هو ان يكون معلم = E_2 .

$$\therefore \quad p_{r=E_1/E_2} = \frac{E_1}{E_2} = \frac{\text{احتمال اجابته نعم}}{\text{احتمال ان يكون معلم}} = \frac{\frac{24}{50}}{\frac{30}{50}} = \frac{24}{50}\,x\,\frac{50}{30} = \frac{24}{30} = \frac{8}{10}$$

مثال (3) : لدينا صندوق يحتوي على (3) كرات بيضاء و (2) كرات سوداء . ليكن (E_1) هو الحادث الاول سحب كرة سوداء , وليكن (E_2) هو الحادث الثاني , سحب كرة سوداء ايضاً. علما بان الكرات المسحوبة لا

تعاد الى الصندوق . الحوادث هنا (E_1) و (E_2) غير مستقلة , اي ان نتيجة السحبة الثانية تعتمد على نتيجة السحبة الاولى .

الحل :

1. احتمال ان تكون السحبة الاولى كرة سوداء = $\frac{2}{5}$ = $\frac{\text{عدد الكرات السوداء}}{\text{عدد الكرات الكلي}}$ = $p_r = (E_1)$

2. احتمال ان تكون السحبة الثانية كرة سوداء بعد ان كانت الكرة الاولى سوداء =

$$p_r = (E_2/E_1) = \frac{\text{عدد الكرات السوداء بعد نقصان واحدة سوداء}}{\text{عدد الكرات الكلي بعد نقصان واحدة سوداء}} = \frac{1}{4}$$

3. احتمال ان تكون كلا الكرتين المسحوبتين سوداء :

$$p_r(E_1\ x\ E_2) = p_r(E_1)\ x\ p_r\ (E_2/E_1) = \frac{2}{5}\ x\ \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

مثال (4) : صندوق يحتوي على (6) كرات حمراء و (4) كرات سوداء , فاذا تم سحب كرتان على التوالي (بدون ارجاع) , ما هو احتمال ان تكون الكرة الثانية حمراء علماً بان الكرة الاولى كانت حمراء ايضاً .

الحل :

بما ان الحوادث غير مستقلة , نطبق القانون التالي :

$$p_r\ (E_1\ E_2) = p_r\ (E_1)\ x\ p_r\ (E_2/E_1) = \frac{5}{15}\ x\ \frac{4}{14} = \frac{4}{42}$$

احتمال ان تكون الكرة الاولى حمراء هو $(\frac{6}{10})$ باعتبار ان عدد الكرات الحمر هو (6) , من مجموع عدد الكرات الكلية وهو (10) . احتمال ان تكون الكرة الثانية حمراء هو $(\frac{5}{9})$, باعتبار انه تم سحب كرة حمراء في المرة الاولى . اذن احتمال ان تكون الكرتان حمراء هو :

$$p_r\ (E_1\ E_2) = p_r\ (E_1)\ x\ p_r\ (E_2/E_1) = \frac{6}{10}\ x\ \frac{5}{9} = \frac{30}{90} = \frac{1}{3}$$

قاعدة ثانية يمكن حل اسئلة الاحتمال الشرطي باستخدام قانون التوافيق والتباديل اذا كانت بالترتيب او بدون ترتيب , فاذا كان لدينا حادثين هما E_1 و E_2 وهما حادثين في فضاء العينة , فان احتمال وقوع الحادث E_1 علماً بان الحادث E_2 قد وقع , ويرمز له بالرمز : $p_r\ (E_2/E_1)$.

$$p_r\,(E_2/E_1) = \frac{p(E_1E_2)}{p(E_1)} = \frac{\text{الحالات المؤاتية لوقوع الحادث } E_2 \text{ و } E_2}{\text{الحالات المؤاتية لوقوع الحادث من عدد الحالات الكلية}}$$

مثال (5) : صندوق يحتوي على (6) كرات حمراء و(4) كرات سوداء , فاذا سحبت منه كرتان على التوالي (بدون ارجاع) , ما هو احتمال ان تكون الكرة الثانية حمراء علماً بان الكرة الاولى كانت حمراء ايضاً ؟

الحل :

a. احتمال اختيار كرتان حمراء هو $nC_r = \binom{n}{r} = \binom{6}{2}$

b. احتمال اختيار كرتان حمراء من الصندوق هو $nC_r = \binom{10}{2}$

حل السؤال بقانون التباديل والتوافيق

$$\therefore\quad p_r\,(E_2/E_1) = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{\frac{6!}{2!\,(6-2)!}}{\frac{10!}{2!\,(10-2)!}} = \frac{1}{3}$$

حل السؤال بقانون الاحتمال الشرطي

$$\therefore\quad p_r\,(E_2/E_1) = \frac{6}{10}\ x\ \frac{5}{9} = \frac{30}{90} = \frac{1}{3}$$

مثال (6) : صُنّف الشباب في احدى المدن كالاتي :

	لديه وظيفة	ليست له وظيفة	المجموع
ذكور	460	40	500
اناث	140	260	400
المجموع	600	300	900

فاذا اخترنا شاب بصورة عشوائية , ما هو احتمال ان يكون ذكر موظف .

الحل :

$$\therefore\quad p_r\,(E_2/E_1) = \frac{p_r\left(\text{احتمال ان يكون ذكر موظف}\right)}{\text{مجموع الموظفين من العدد الكلي}}\ x\ \frac{\frac{460}{900}}{\frac{600}{900}} = \frac{460}{900}\ x\ \frac{900}{600} = \frac{460}{600} = \frac{23}{30}$$

الفصل الرابع عشر

التوزيع البوسواني Poisson Distribution

في بعض التجارب قد يحدث المتغير العشوائي في جزء من وقت محدد , مثلا عدد المكالمات التلفونية في الساعة المستعملة من قبل دائرة ما , أو عدد ايام تعطيل المدارس بسبب الامطار مثلاً , أو عدد الاخطاء في صفحة مطبوعة ..... الخ. هذه التجارب تسمى تجارب بوسوان ومن صفاتها :

1. متوسط عدد ظهور النجاحات = μ .
2. احتمال حصول نجاح واحد في فترة قصيرة يتناسب مع طول الوقت .
3. احتمال حصول اكثر من نجاح واحد في مثل هذه الفترة القصيرة هو احتمال نادر .

يعرف التوزيع الاحتمالي للمتغير العشوائي البوسواني (y) , الذي يمثل عدد النجاحات التي تحدث في فترة محددة هو:

$$p(y) = \frac{e^{-\mu}}{y!}$$

حيث ان :

y = 1 , 2 , 3, ……etc.

μ = متوسط عدد النجاحات

e = 2.71828 (constant)

مثال : اذا كان متوسط عدد الايام التي تعطل فيها الدراسة في مدرسة معينة بسبب سقوط الثلوج في فصل الشتاء هو (4) ايام , ما هو احتمال ان المدارس في هذه المدينة ستعطل فيها الدراسة لمدة (6) ايام خلال الشتاء .

الحل :

$$p(y) = \frac{e^{-\mu}}{y!} = \frac{(2.71828)^{-4}\ x\ (4)^6}{6x5x4x3x2x1} = 0.1024$$

الباب الثالث

معامل الارتباط

ومخطط الانتشار

الفصل الخامس عشر

معامل الارتباط Correlation Coefficient

يعرف الارتباط بانه العلاقة الرياضية التي تربط بين متغيرين $(x\,,y)$ أو اكثر , أو هو قوة العلاقة بين متغيرين , وهي قيمة حقيقية كمية خالية من الوحدات (وحدات القياس) , ويرمز له بالرمز (r) . وهو احد انواع العلاقات بين المتغير التابع والمتغير المستقل , حيث ان (X) يعتبر متغير مستقل و (y) متغير تابع , ويعني ذلك اذا تغير احد المتغيرين باتجاه معين فان المتغير الاخر يميل الى التغير باتجاه معين ايضاً, بمعنى انه اذا تغير احد المتغيرين فان المتغير الآخر قد يتبعه. إما ان يتبعه في نفس الاتجاه فيكون الارتباط طردي , او يتبعه في عكس الاتجاه فيكون الارتباط عكسي, ويقال ان المتغيرين مستقلين عندما ينعدم الارتباط. درجة الارتباط تقاس بعدد يتراوح مقداره بين $(1\ ,\ 0\ ,\ -1)$. الارتباط ليس شرطاً ان يتغير احد المتغيرين دائماً بتغير الاخر. وبناء على هذه الدرجة او المقياس , يتم تصنيف درجة او علاقة الارتباط الى ثلاث اصناف :

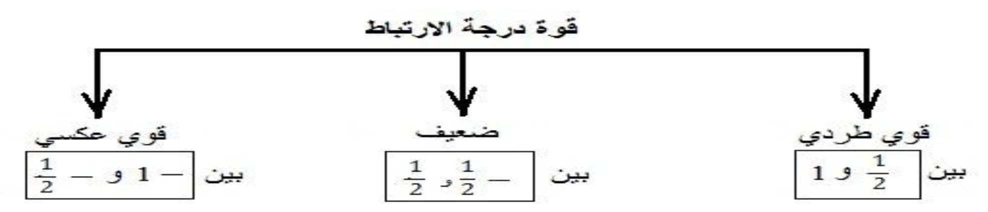
1. درجة الارتباط = $(+1)$: وهذا يعني ان درجة الارتباط (موجب) , ويصنف الارتباط على انه (ارتباط طردي تام) .
2. درجة الارتباط = (-1) : وهذا يعني ان درجة الارتباط (سالب) , ويصنف الارتباط على انه (ارتباط عكسي تام) .
3. درجة الارتباط = (صفر) : وهذا يعني انه لا يوجد ارتباط .



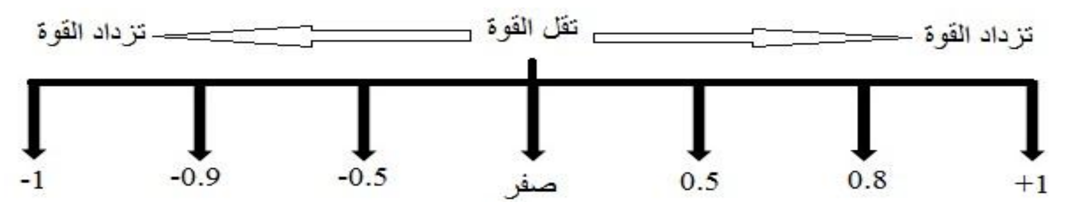
اما قوة الارتباط, يمكن ان تصنف الى ثلاثة اصناف وكما يلي :

1. اذا كانت درجة الارتباط تقع بين $\left(\frac{1}{2}\right.$ و $\left.\mathbf{1}\right)$, فهذا يعني ان الارتباط قوي .
2. اذا كانت درجة الارتباط تقع بين $\left(-\frac{1}{2}\right.$ و $\left.\frac{1}{2}\right)$, فهذا يعني ان قوة الارتباط تصنف على انها ضعيفة .
3. اذا كانت درجة الارتباط تقع بين $\left(-\frac{1}{2}\right.$ و $\left.-\mathbf{1}\right)$ فان قوة درجة الارتباط تصنف على انها عكسي قوي.

وكما هو موضح في المخطط التالي :



علماً بان قوة درجة الارتباط تزداد كلما اقتربنا من الاطراف سواء كانت باتجاه الطرف القوي الطردي أو باتجاه الطرف القوي العكسي , وتقل كلما ابتعدنا عن الاطراف باتجاه المركز , اي باتجاه الصفر , كما في المخطط التالي :



مثال : عين اي من الارقام التالية يمثل معامل الارتباط الاقوى : 0.3 , -0.9 , -0.5 , 0.6

الحل : أقرب رقم الى الاطراف سواء كان باتجاه طرف القيم الطردية أو طرف القيم العكسية (1) أو (-1) , فهو الرقم (-0.9) .

<u>طرق قياس معمل الارتباط</u>

<u>معمل الارتباط :</u> هو المقياس الرقمي أو القيمة العددية لقوة الارتباط بين متغيرين مثل (X) و (y) , ويمتاز معامل الارتباط بمجموعة من الخصائص وهي :

1. تزداد قوة العلاقة كلما اقترب معامل الارتباط من الاطراف , أي (-1 و $+1$) ويسمى :
   a. عندما $r = +1$ تسمى ارتباط طردي تام .
   b. عندما $r = -1$ تسمى ارتباط عكسي تام .
2. تقل العلاقة كلما اقترب معامل الارتباط من المركز , اي باتجاه القيمة صفر , ويسمى :
   a. عندما $r = 0$ لا يوجد ارتباط .
3. توصف علاقة معامل الارتباط عندما تقع القيمة بين الاطراف ومركز التوزيع وتوصف كما يلي :

a. عندما تقع قيمة معامل الارتباط بين (0 و 1) , تسمى العلاقة في هذه الحالة ارتباط طردي موجب .

b. عندما تقع قيمة معامل الارتباط بين (-1 و 0) , تسمى العلاقة في هذه الحالة ارتباط عكسي سالب .

طرق قياس معامل الارتباط وقوته :

تقسم طرق قياس معامل الارتباط وقوته الى قسمين :

1. معامل ارتباط بيرسون Pearson’s Correlation
2. معامل ارتباط سبيرمان Spearman Correlation

معامل ارتباط بيرسون Pearson’s Correlation

ويسمى كذلك معامل ارتباط العزوم , وهو المقياس الاقوى لانه تعامل مع نفس القيم للمتغيرات الداخلة في الحساب .ويتم حسابه كما يلي :

اذا كان لدينا عدد (n) من المتغيرات (X) أو (y) (متغيرات من ازواج القيم) التي تمثل مثلاً (وزن الطالب, طول الطالب) أو (تركيز عنصر الحديد , تركيز عنصر النحاس) أو مثلاً (درجة الطالب في مادة الرياضيات مع درجة الطالب في مادة الاحصاء) وهكذا , ويرمز لها كما يلي :

$$(X_1 y_1)\,,(X_2 y_2)\,,(X_3 y_3)\,, \dots \dots \dots ,(X_n y_n)$$

معامل الارتباط يحسب باحدى العلاقات التالية :

1. $$r = \frac{\sum (X-\bar{X})(y-\bar{y})}{n \; x \; S_x \; x \; S_y}$$

2. $$r = \frac{\sum Xy - n\,\bar{X}\,\bar{y}}{\sqrt{\sum X^2 - n \; x \; (\bar{X})^2} \; x \; \sqrt{\sum y^2 - n \; x \; (\bar{y})^2}}$$

3. $$r = \frac{\frac{\sum x\,y}{n} - \bar{x}\;\bar{y}}{S_x S_y}$$

حيث ان :

n = عدد القيم (X) أو (y) التي تمثل (أزواج القيم)

$\bar{X}$ = الوسط الحسابي لقيم الظاهرة (X)

X = قيم الظاهرة X

y = قيم الظاهرة y

$\bar{y}$= الوسط الحسابي لقيم الظاهرة y

S_X= X الانحراف المعياري للظاهرة

S_y= y الانحراف المعياري للظاهرة

لحساب معامل الارتباط نتبع الخطوات التالية :

1. ايجاد (n) وهو عدد ازواج القيم من X و y .
2. ايجاد الوسط الحسابي للظاهرتين X و y .
3. ايجاد الانحراف المعياري للظاهرتين X و y .
4. ايجاد مجموع حاصل ضرب كل من الظاهرتين $\sum X\, x\, y$ أو $\sum(X-\bar{X})\,(y-\bar{y})$.
5. نطبق احدى العلاقتين للحصول على قيمة معامل ارتباط بيرسون .

مثال (1): البيانات الدرجة في الجدول التالي تمثل تركيز عنصر الحديد (Fe) مع تركيز عنصر الخارصين (Zn) ,

في تسعة نماذج مستحصلة من رواسب جداول تصريف المياه السطحية . احسب معامل الارتباط بيرسون بين تركيز عنصري الحديد والخارصين.

Fe %	2	2	5	4	5	6	3	5	4
Zn %	3	5	7	8	9	11	6	8	6

الحل :

1. نوجد الوسط الحسابي للمتغيرين بموجب العلاقات التالية :
2.

$$\bar{X} = \frac{\sum Fe\,\%}{n} = \frac{36}{9} = 4$$

$$\bar{y} = \frac{\sum Zn\,\%}{n} = \frac{63}{9} = 7$$

3. يتم عمل جدول لغرض تسهيل مهمة ايجاد قيم كل من : $\sum(X-\bar{X})\,(y-\bar{y})$ وكما يلي:

Fe %	Zn %	$Fe - \overline{Fe}$	$(Fe - \overline{Fe})^2$	$Zn - \overline{Zn}$	$(Zn - \overline{Zn})^2$	$(Fe - \overline{Fe})\ x\ (Zn - \overline{Zn})$
2	3	-2	4	- 4	16	8
2	5	-2	4	- 2	4	4
5	7	1	1	0	0	0
4	8	0	0	1	1	0
5	9	1	1	2	4	2
6	11	2	4	4	16	8
3	6	-1	1	- 1	1	1
5	8	1	1	1	1	1
4	6	0	0	- 1	1	0
المجموع			$\sum 16$		$\sum 44$	$\sum 24$

4. نوجد قيمة الانحراف المعياري لعنصري الحديد والزنك من العلاقات التالية :

1. $$S_X = \sqrt{\frac{\sum(Fe-\overline{Fe})^2}{n}} \quad or \quad S_X = \sqrt{\frac{\sum(X)^2}{n} - (\overline{X})^2} = \sqrt{\frac{16}{9}} = \frac{4}{3}$$

2. $$S_y = \sqrt{\frac{\sum(Zn-\overline{Zn})^2}{n}} \quad or \quad S_y = \sqrt{\frac{\sum(y)^2}{n} - (\overline{y})^2} = \sqrt{\frac{44}{9}} = \frac{\sqrt{44}}{3}$$

5. نطبق احد القوانين الخاصة بايجاد قيمة معامل الارتباط بيرسون :

$$r = \frac{\sum(X - \overline{X})(y - \overline{y})}{n\ x\ S_x\ x\ S_y} = \frac{24}{9\ x\ \frac{4}{3}\ x\ \frac{\sqrt{44}}{3}} = \frac{24}{\sqrt[4]{44}} = 0.0905$$

الارتباط موجب , وهذا يعني انه كلما ازداد تركيز عنصر الحديد ازداد معه تركيز عنصر الخارصين.

مثال (2) : أوجد معامل ارتباط بيرسون للمتغيرين (X) و (y) المذكورة في الجدول التالي :

X	1	2	3	4	5
y	1	− 1	− 4	− 6	− 5

الحل:

معامل ارتباط بيرسون يعطى بالقانون التالي : $r = \frac{\sum(X-\bar{X})(y-\bar{y})}{n\ x\ S_x\ x\ S_y}$

الوسط الحسابي للمتغير (X) = $\bar{X} = \frac{\sum X}{n} = \frac{1+2+3+4+5}{5} = +3$

الوسط الحسابي للمتغير (y) = $\bar{y} = \frac{\sum y}{n} = \frac{1+(-1)+(-4)+(-6)+(-5)}{5} = -3$

الانحراف المعياري للمتغير (X) = $S_X = \sqrt{\frac{\sum(X)^2}{n} - (\bar{X})^2}$

الانحراف المعياري للمتغير (y) = $S_y = \sqrt{\frac{\sum(y)^2}{n} - (\bar{y})^2}$

لغرض تسهيل عملية اجراء الحسابات نلجأ الى تنظيم البيانات كما في الجدول التالي :

X	y	$(X - \bar{X})$	$(y - \bar{y})$	$(X - \bar{X})\ x\ (X - \bar{X})$	$(X)^2$	$(y)^2$
1	1	1-3 = -2	$1-(-3) = 4$	$-2\ x\ 4 = -8$	1	1
2	-1	2-3 = -1	$-1-(-3) = 2$	$-1\ x\ 2 = -2$	4	1
3	-4	3-3 = 0	$-4-(-3) = -1$	$0\ x\ -1 = 0$	9	16
4	-6	4-3 = 1	$-6-(-3) = -3$	$1\ x\ -3 = -3$	16	36
5	-5	5-3 = 2	$-5-(-3) = -2$	$2\ x\ -2 = -2$	25	25
$\sum 15$	$\sum -15$				$\sum 55$	$\sum 79$

الانحراف المعياري للمتغير (X)=

$$S_X = \sqrt{\frac{\sum(X)^2}{n} - (\bar{X})^2} = \sqrt{\frac{55}{5} - (3)^2} = \sqrt{11-9} = \sqrt{2}$$

الانحراف المعياري للمتغير (y)=

$$S_y = \sqrt{\frac{\sum (y)^2}{n} - (\bar{y})^2} = \sqrt{\frac{79}{5} - (-3)^2} = \sqrt{15.8 - 9} = \sqrt{6.8}$$

نستخرج معامل ارتباط بيرسون حسب القانون الاول :

$$r = \frac{\sum (X - \bar{X})(y - \bar{y})}{n\ x\ S_x\ x\ S_y} = \frac{-17}{5\ x\ \sqrt{2}\ x\ \sqrt{6.8}} = \frac{-17}{18.4} = -0.92$$

اذن الارتباط هو من نوع (ارتباط عكسي قوي)

نستخرج معامل ارتباط بيرسون حسب القانون الثاني وهو :

$$r = \frac{\sum Xy - n\,\bar{X}\,\bar{y}}{\sqrt{\sum X^2 - n\ x\ (\bar{X})^2}\ x\ \sqrt{\sum y^2 - n\ x\ (\bar{y})^2}}$$

نستخرج المتغيرات المطلوبة بواسطة عمل جدول لتسهيل مهمة اجراء الحسابات وكما يلي :

X	y	$X\ x\ y$	X^2	y^2
1	1	1	1	1
2	−1	−2	4	1
3	−4	− 12	9	16
4	−6	− 24	16	36
5	−5	− 25	25	25
		$\sum -62$	$\sum 55$	$\sum 79$

$$r = \frac{\sum Xy - n\,\bar{X}\,\bar{y}}{\sqrt{\sum X^2 - n\ x\ (\bar{X})^2}\ x\ \sqrt{\sum y^2 - n\ x\ (\bar{y})^2}} =$$

$$\frac{(-62) - [5\ x\ 3\ x\ (-3)]}{\sqrt{55 - [5\ x\ (3)^2]}\ x\ \sqrt{79 - [5\ x\ (-3)^2]}} = \frac{(-62) - (-45)}{\sqrt{55 - 45}\ x\ \sqrt{79 - 45}} =$$

$$\frac{(-17)}{\sqrt{10}\ x\ \sqrt{34}} = \frac{-17}{18.4} = -0.92$$

اذن نوع الارتباط هو ارتباط (عكسي قوي)

مثال (3):

لدراسة علاقة الصادرات بالميزان التجاري خلال عدة سنوات, تم الحصول على عشرة قراءات تقريبية لقيمة صادرات جمهورية العراق التي يرمز لها بالرمز (x) وقيمة الميزان التجاري الذي يرمز له بالرمز (y) بملايين الدنانير , وكما موضح في الجدول التالي:

y	1	3	8	7	6	5	7	8	12	12
X	9	11	17	18	19	16	16	19	23	23

هل توجد علاقة ارتباط خطية؟ ما نوعها ؟ وما مدى قوتها؟

الحـــل:

A. نكون جدول يتم فيه حساب المتغيرات المطلوبة لغرض تعويضها في قانون ارتباط بيرسون وكما يلي:

X	y	xy	x^2	y^2
9	1	9	81	1
11	3	33	121	9
17	8	136	289	64
18	7	126	324	49
19	6	114	361	36
16	5	80	256	25
16	7	112	256	49
19	8	152	361	64
23	12	276	529	144
23	12	276	529	144
$\sum x = 171$	$\sum y = 69$	$\sum xy = 1314$	$\sum x^2 = 3107$	$\sum y^2 = 585$

B. نحسب الوسط الحسابي لكلا المتغيرين x, y: $\bar{x} = \frac{\sum x}{n} = \frac{171}{10} = 17.1 \quad , \ \bar{y} = \frac{\sum y}{n} = \frac{69}{10} = 6.9$

C. نحسب الانحراف المعياري لكلا المتغيرين x, y:

$$S_X = \sqrt{\frac{\sum(X)^2}{n} - (\bar{X})^2} = \sqrt{\frac{3107}{10} - (17.1)^2} = \sqrt{310.7 - 292.41} = \sqrt{18.29}$$

$$S_y = \sqrt{\frac{\sum(y)^2}{n} - (\bar{y})^2} = \sqrt{\frac{585}{10} - (6.9)^2} = \sqrt{58.5 - 47.61} = \sqrt{10.89}$$

D. نستخرج معامل ارتباط بيرسون من العلاقة التالية:

$$r = \frac{\frac{\sum x\,y}{n} - \bar{x}\;\bar{y}}{S_x S_y} = r = \frac{\frac{1314}{10} - (17.1)(6.9)}{\sqrt{18.2}\sqrt{10.89}} = \frac{13.41}{14.11} = 0.95$$

العلاقة هي ارتباط طردي قوي , 0.95

مثال (4):

تم تسجيل ستة قراءات تقريبية لحجم الانتاج وحجم صادرات النفط الخام للعراق (مليار برميل) خلال ست سنوات وكانت كما يلي:

حجم الصادرات (y)	2	2	2	1	1	1
حجم الانتاج (x)	3	4	2	2	2	2

هل توجد علاقة ارتباط خطية بين حجم الانتاج وحجم صادرات النفط الخام؟

الحـــل:

A. نكون جدول يتم فيه حساب المتغيرات المطلوبة لغرض تعويضها في قانون ارتباط بيرسون وكما يلي:

X	y	xy	x^2	y^2
3	2	6	9	4
4	2	8	16	4
2	2	4	4	4
2	1	2	4	1
2	1	2	4	1
2	1	2	4	1
$\sum x = 15$	$\sum y = 9$	$\sum xy = 24$	$\sum x^2 = 41$	$\sum y^2 = 15$

B. نحسب الوسط الحسابي لكلا المتغيرين x ,y: $\bar{x} = \frac{\sum x}{n} = \frac{15}{6} = 2.5$, $\bar{y} = \frac{\sum y}{n} = \frac{9}{6} = 1.5$

C. نحسب الانحراف المعياري لكلا المتغيرين x ,y:

$$S_X = \sqrt{\frac{\sum(X)^2}{n} - (\bar{X})^2} = \sqrt{\frac{41}{6} - (2.5)^2} = \sqrt{6.83 - 6.25} = \sqrt{0.58}$$

$$S_y = \sqrt{\frac{\sum(y)^2}{n} - (\bar{y})^2} = \sqrt{\frac{15}{6} - (1.5)^2} = \sqrt{2.5 - 2.25} = \sqrt{0.25}$$

D. نستخرج معامل ارتباط بيرسون من العلاقة التالية:

$$r = \frac{\frac{\sum x\,y}{n} - \bar{x}\ \bar{y}}{S_x S_y} = r = \frac{\frac{24}{6} - (1.5)(2.5)}{\sqrt{0.58}\sqrt{0.25}} = \frac{0.25}{0.38} = 0.66\,,$$ العلاقة هي ارتباط طردي

معمل الارتباط سبيرمان Spearman Coefficient :

ويسمى كذلك معامل ارتباط الرتب. ويستخدم اذا كانت الارقام كبيرة , أو اذا لم تتواجد القيم الحقيقية للمتغيرات ولكن تتوفر لدينا ترتيب لهذه القيم , اي يكون كلا المتغيرين ترتيبيين, او احدهما ترتيبي والاخر كمي. يتم حساب معامل ارتباط سبيرمان حسب العلاقة التالية :

$$\boldsymbol{r = 1 - \frac{6\ x\ \sum F^2}{n\ x\ (n^2\ - 1)}}$$

حيث ان :

$\boldsymbol{r} =$ معامل ارتباط سبيرمان

$F =$ (y رتبة و X رتبة ,وتسمى كذلك رتبة المتغيرات) y و X ترتيب كل متغير من المتغيرات

n = y أو X عدد قيم المتغير

يتم استخدام هذه الصيغة الرياضية عندما تكون قيم الوسط الحسابي للمتغيرات $\boldsymbol{\bar{X}}$ و $\boldsymbol{\bar{y}}$ كبيرة , أو اذا لم يتم اعطاء القيم للمتغيرات $\boldsymbol{X}$ و $\boldsymbol{y}$ وتم اعطاء ترتيبها , بعد ان يتم ترتيبها تنازلياً .

مثال (1) : أوجد معامل ارتباط سبيرمان للمتغيرين (X & y) , والبيانات معطاة في الجدول التالي :

9	11	5	13	12	4	6	10	8	X
150	160	120	180	165	130	150	160	150	y

الحل :

نتبع الصيغة الرياضية التالية لايجاد معامل ارتباط سبيرمان :

$$\boldsymbol{r = 1 - \frac{6\ x\ \sum F^2}{n\ x\ (n^2\ - 1)}}$$

n = عدد النتائج او عدد المفردات

F = $\left(\boldsymbol{y}\right.$ رتبة $-$ $\left.\boldsymbol{X}\right)$ رتبة

يتم عمل جدول لغرض تسهيل مهمة ايجاد الرتب للمتغيرات :

قيم X	8	10	6	4	12	13	5	11	9
نرتب القيم تنازليا	13	12	11	10	9	8	6	5	4
نرقم القيم	1	2	3	4	5	6	7	8	9
رتبة X	1	2	3	4	5	6	7	8	9

قيم y	150	160	150	130	165	180	120	160	150
نرتب القيم تنازليا	180	165	160	160	150	150	150	130	120
نرقم القيم	1	2	3	4	5	6	7	8	9
رتبة y	1	2	$(\frac{4+3}{2})$ = 3.5	$(\frac{4+3}{2})$ = 3.5	$(\frac{7+6+5}{3})$ = 6	$(\frac{7+6+5}{3})$ = 6	$(\frac{7+6+5}{3})$ = 6	8	9

ملاحظة: عند ترتيب القيم تنازليا , في حالة وجود قيم متشابهة أو اعداد متشابهة , في هذه الحالة يتم احتساب رتبة المتغير باخذ معدل رقم القيم المتشابهة واعتمادها كقيمة لرتبة المتغير, وكما تم احتسابها في الجدول اعلاه.

يتم عمل جدول ثاني لغرض استخراج المتغيرات التي تدخل في الحسابات للصيغة الرياضية وكما يلي:

X	y	رتبة X	رتبة y	**F** = (رتبة **y** − رتبة **X**)	F^2
8	150	6	6	صفر	صفر
10	160	4	3.5	0.5	0.25
6	150	7	6	1	1
4	130	9	8	1	1
12	165	2	2	صفر	صفر
13	180	1	1	صفر	صفر
5	120	8	9	- 1	1
11	160	3	3.5	- 0.5	0.25
9	150	5	6	- 1	1
					4.5

اذن معامل ارتباط سبيرمان =

$$r = 1 - \frac{6\ x\ \sum F^2}{n\ x\ (n^2\ - 1)} = 1 - \frac{6\ x\ 4.5}{9\ x\ (81 - 1)} = 1 - \frac{27}{720} = 0.963$$

اذن الارتباط هو من نوع الارتباط (طردي قوي)

مثال (2): البيانات التالية تمثل درجات ستة طلاب في مادتي الاحصاء والرياضيات . أوجد معامل الارتباط بين تلك المادتين ؟

الاحصاء	ممتاز	جيدجداً	جيد	جيد	ضعيف	مقبول
الرياضيات	مقبول	ضعيف	مقبول	جيد	جيدجداً	ممتاز

الحل :

بما انه لدينا في هذا المثال رتب او تقديرات ولا توجد لدينا درجات أو قيم للطلاب , اذن لا بد من استخدام معامل ارتباط سبيرمان لانه المعامل الذي يختص بالرتب في حالة عدم وجود قيم. يتم عمل جدول يتم فيه ترتيب التقديرات تنازلياً وثم نستخرج المتغيرات التي تدخل في حل المعادلة وكما يلي :

الاحصاء	رتبة الاحصاء	الرياضيات	رتبة الرياضيات	$F =$ رتبة الحصاء – رتبة الرياضيات	$(F)^2$
ممتاز	1	ممتاز	1	صفر	صفر
جيدجداً	2	جيدجداً	2	صفر	صفر
جيد	$\frac{3+4}{2} = 3.5$	جيد	3	0.5	0.25
جيد	$\frac{3+4}{2} = 3.5$	مقبول	$\frac{4+5}{2} = 4.5$	- 1	1
مقبول	5	مقبول	$\frac{4+5}{2} = 4.5$	0.5	0.25
ضعيف	6	ضعيف	6	صفر	صفر
					$\sum F^2 = 1.5$

اذن معامل سبيرمان =

$$r = 1 - \frac{6\ x\ \sum F^2}{n\ x\ (n^2\ - 1)} = 1 - \frac{6\ x\ 1.5}{6\ x\ ((6)^2 - 1)} = 1 - \frac{9}{6\ x\ 35} = 0.957$$

وهذا يعني ان بين المادتين (الاحصاء والرياضيات) ارتباط طردي قوي .

مثال اثرائي (3): أوجد معامل ارتباط سبيرمان ومعامل ارتباط بيرسون لقيم (X & y) المدرجة في الجدول التالي وبين نوع الارتباط .

3	4	1	3	8	5	X
4	5	1	4	10	6	y

الحل: معامل ارتباط سبيرمان = 1 , معامل ارتباط بيرسون = 0.997 ≅ 1 .

الفصل السادس عشر

مخطط الانتشار وخط الانحدار

Scatter Diagram and Regression Line

مخطط الانتشار Scatter Diagram :

وهو التمثيل البياني للعلاقة بين اي متغيرين الذي يقيس العلاقة بين متغيرين كميين, ويكون ذلك بواسطة اسقاط نقاط المتغيرين على المحورين الافقي والعمودي .

خط الانحدار Regression Line :

وهو العلاقة الخطية التي تربط بين اي متغيرين ومن خلالها نستطيع تقدير قيمة المتغير الاول من المتغير الثاني باستخدام معادلة خط الانحدار والتي تسمى كذلك معادلة الخط المستقيم وهي:

$$y = a + b \ x \ X$$

حيث ان :

X & y = متغيران

a & b = ثوابت

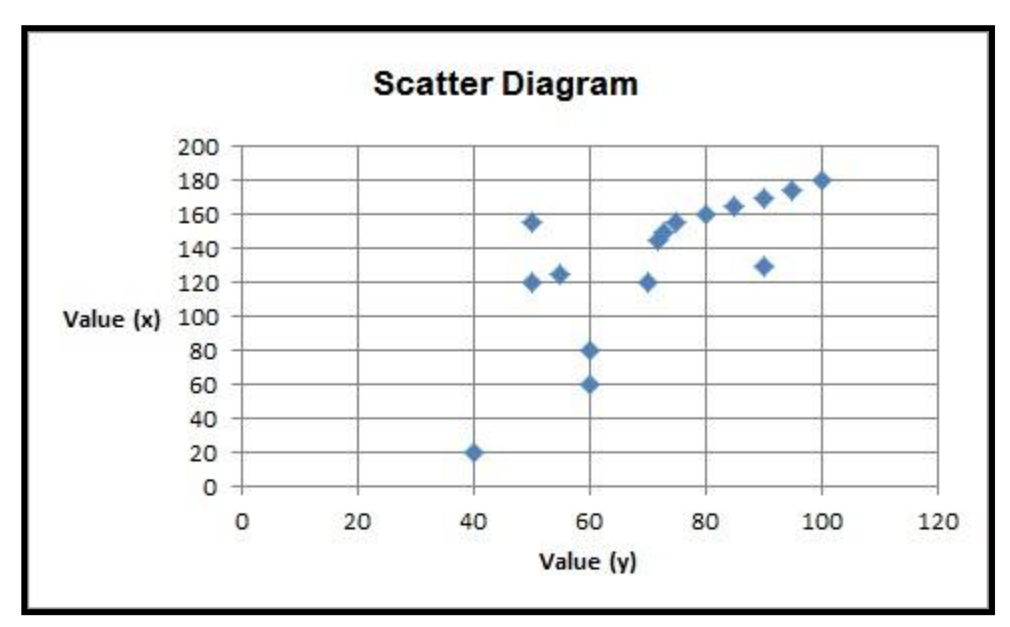
1. مخطط الانتشار :

نوع من انواع عرض وتحليل البيانات في حالة وجود توزيع لمتغيرين او اكثر . وهو عبارة عن عمل انتشار لنقاط المتغيرين على المحورين (X & y) , ومن خلال شكل توزيع هذه النقاط أو شكل العلاقة الناتجة بين المتغيرين ويسمى في كثير من الاحيان بالشكل الانتشاري , يمكن تكوين فكرة جيدة عما اذا كان هناك علاقة ارتباط بين المتغيرين (نوجد علاقة بينهما) أو لا توجد .

في مخطط الانتشار سوف نحصل على ازواج من القيم , مثلا في المتغير الاول (X) , اذا كان لدينا عدة قيم وهي $(X_1 \ , \ X_2 \ , \ X_3 \ , \ldots \ldots , X_n \ )$, بالمقابل سوف توجد لدينا قيم للمتغير الثاني (y) , وهي $(y_1 \ , \ y_2 \ , \ y_3 \ , \ \ldots .. , \ y_n)$, وهذا يعني ان لكل قيمة للمتغير (X) توجد قيمة مقابلة لها للمتغير (y) وبذلك سوف نحصل على ازواج القيم والتي تكتب بالشكل التالي:

$$(X_1 \ , y_1 \ ) \ , (X_2 \ , y_2) \ , (X_3 \ , X_3) \ , \ldots \ldots , (X_n \ , X_n)$$

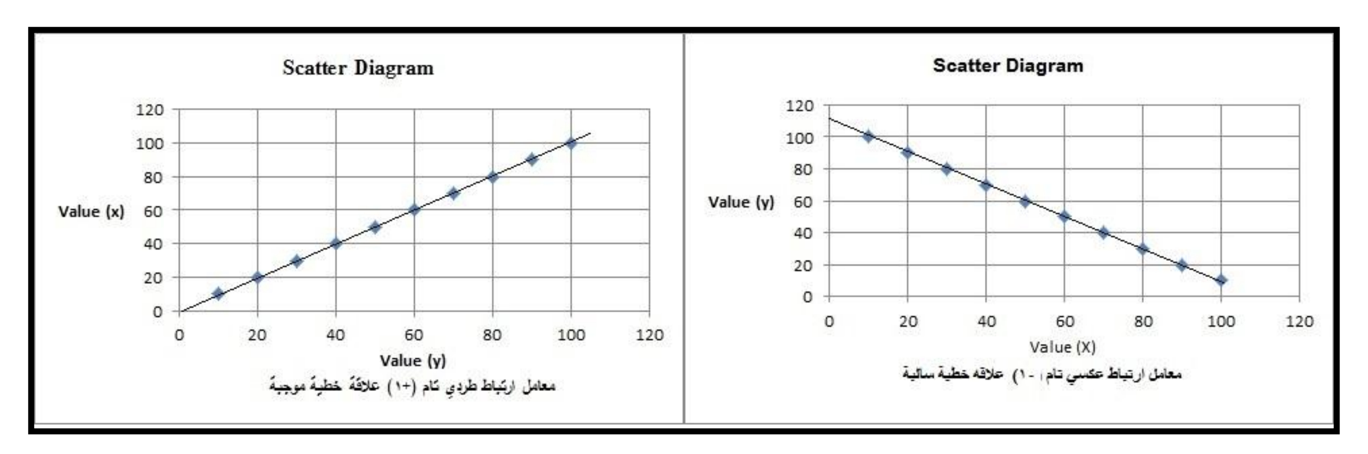
بهذه الحالة فهي تشبه الاحداثيات التي يمكن تمثيلها على المحور الافقي والمحور العمودي , حيث يمكن تمثيل كل زوج من هذه القيم على شكل نقطة في مخطط الانتشار , وهكذا بالنسبة لبقية ازواج القيم حيث تسقط جميعاً على مخطط الانتشار لنحصل في النهاية على عدة نقاط تمثل شكل العلاقة البيانية التي تربط بين المتغيرين $(y \ , X)$ والتي نستنتج منه قوة العلاقة ونوع الارتباط بين المتغيرين , ومثال على ذلك مخطط الانتشار التالي:



الاستنتاج الذي يمكن ان نحصل عليه من الشكل اعلاه هو ان نقول ان النقاط التي تمثل القيم العالية توجد بينهما علاقة , بينما لا توجد علاقة في النقاط التي تمتلك قيم واطئة , وتوجد بعض النقاط بينهما علاقة عشوائية.

اذا كانت نقاط المخطط الانتشاري متقاربة مع بعضها البعض أو يمكن ان تقع على خط مستقيم , فاننا بهذه الحالة نتوقع وجود علاقة ارتباط جيدة بين المتغيرين سواء كانت هذه العلاقة طردية ام عكسية , اما اذا كانت النقاط متباعدة فننا نتوقع ان الارتباط ضعيف او لا يوجد ارتباط .

الاشكال التاية توضح انواع شكل مخطط الانتشار وكما في الشكل (1−5) و (2−5) :



شكل (5-1)

Scatter Diagram
معامل ارتباط طردي موجب

Scatter Diagram
ارتباط عكسي سالب

Scatter Diagram
لا توجد علاقة ارتباط -- علاقة لا خطية

شكل (5-2)

نلاحظ انه في حالة الارتباط الطردي التام , ان قيم المتغير الاول (X) تزداد طردياً بنفس المقدار في زيادة قيم المتغير (y) , أما في حالة الارتباط العكسي , ان قيم المتغير الاول تزداد مع نقصان في قيم المتغير الثاني, والعكس بالعكس. في حالة العلاقة الطردية الموجبة , ان قيم المتغير الاول تزداد ولكن ليس بنفس مقدار زيادة قيم

المتغير الثاني وبالعكس. في حالة العلاقة العكسية السالبة , نلاحظ ان قيم المتغير الاول تزداد مع نقصان غير متوافق وبمقدار معين في قيم المتغير الثاني . في حالة مخطط الانتشار العشوائي , أو في حالة عدم وجود علاقة ارتباط , ان قيم المتغير الاول ليس لها علاقة في حالة حصول اية زيادة او نقصان في قيم المتغير الثاني, أي لا توجد علاقة اقتران بين المتغيرات وكل متغير له قيم مستقلة عن الاخرى .

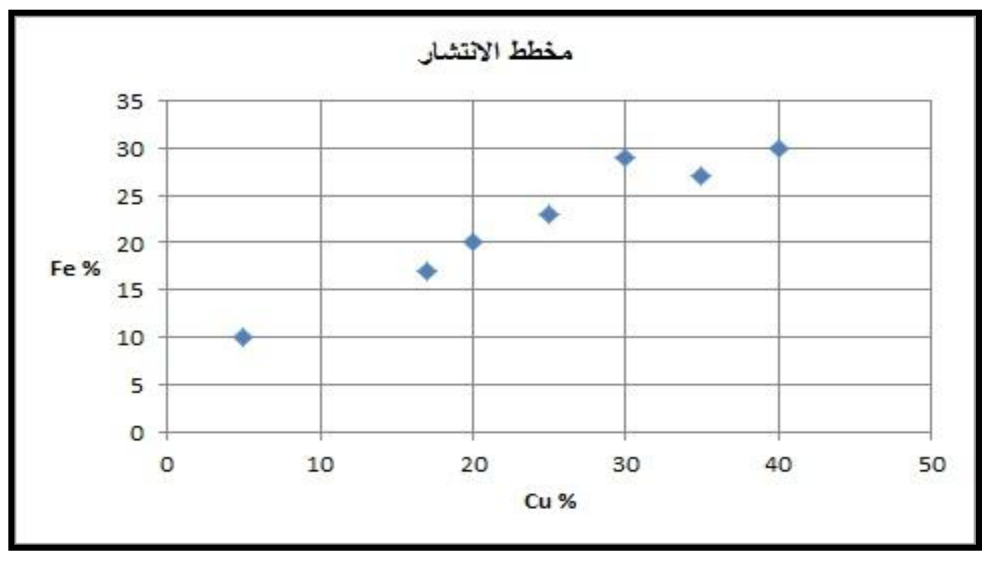
يستخدم مخطط الانتشار في معالجة البيانات الاحصائية بصورة واسعة كونه سهل الاستخدام ويتم الحصول منه على معلومات مهمة وبصورة سريعة , يمكن من خلاله ان نستدل او نستنتج قيم المتغير الثاني بدلالة المتغير الاول وبالعكس .

مثال على ذلك : اثناء جمع عينات من جداول تصريف الانهار لاغراض تحليل العناصر الاثرية (Trace Elements) مثل (Zn , Fe , Ca , Pb , Mn , ….,etc) , اذا تواجدت بعض من هذه العناصر بتراكيز عالية , ممكن للعناصر الاخرى ان تتواجد كذلك بتراكيز عالية ايضاً اذا كانت بينهما علاقة اقتران او علاقة طردية , وممكن ان تكون بالعكس او لا توجد بينهما اي علاقة من حيث قيم تراكيزها في جداول تصريف الانهار. من خلال دراسة هذه العلاقة بين العناصر الاثرية , ممكن ان نستدل منها على طبيعة البيئة الترسيبية لها واسباب هذا التواجد وهل ان تواجدها كان في نفس بيئة الترسيب ام ان هناك اسباب اخرى ادت الى زيادة تراكيز بعض من هذه العناصر او بالعكس .

مثال (1): الجدول التالي يمثل تراكيز عنصري الحديد والنحاس مقاسة بالنسبة المئوية . ارسم مخطط الانتشار وبين نوع العلاقة بين العنصرين ؟

Fe %	Cu %
10	5
17	17
20	20
23	25
29	30
27	35
30	40

الحل :



شكل (3−5)

نلاحظ من طبيعة انتشار النقاط , ان هناك علاقة طردية موجبة بين العنصرين .

2. خط الانحدار :

لمعرفة طبيعة العلاقة بين اي متغيرين , نرسم شكل الانتشار كما تم ذكره سابقاً , ومن شكل الانتشار نلاحظ مدى تباعد أو تجمع النقاط حول خط مستقيم , فاذا كانت النقاط تتجمع حول خط مستقيم فاننا نقول ان ان العلاقة بين المتغيرين (X , y) هي علاقة خطية . اذا كانت جميع النقاط تقع على خط مستقيم , بهذه الحالة يكون معامل الارتباط يساوي (1 +) .

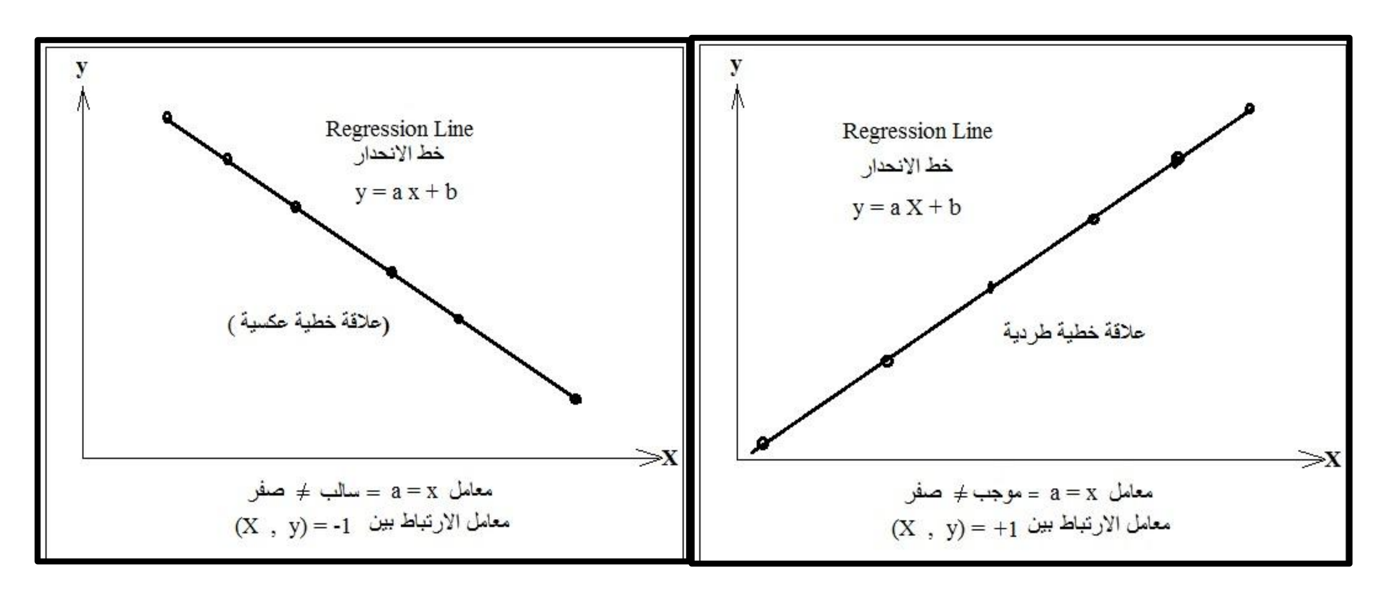
اذا كان معامل الارتباط بين المتغيرين طردي تام او عكسي تام , فان العلاقة الرياضية بين المتغيرين (X , Y) هي :

$$y = aX + b \qquad or \qquad X = cy + d$$

حيث ان : a , b , c, d هي اعداد حقيقية وبشرط ان :

$a \neq 0$ صفر و $c \neq 0$ صفر

الشكل التالي يوضح العلاقة الخطية بين المتغيرين (X , y) سواء كانت العلاقة يمثل ارتباط طردي تام تساوي (+1) أو ارتباط عكسي تام يساوي (−1) . شكل رقم (4−5) .



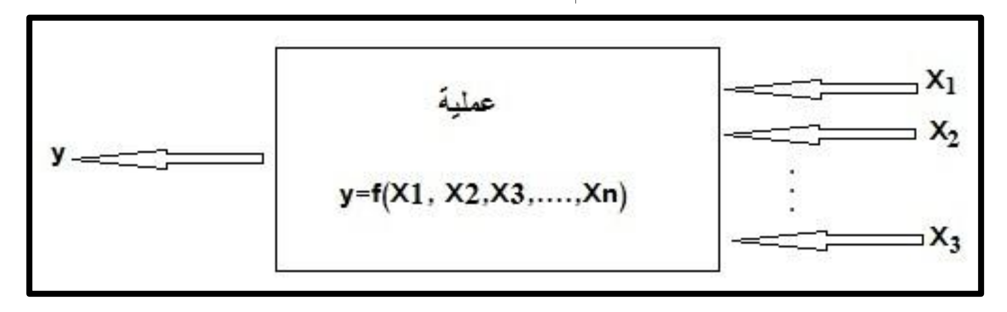
شكل رقم (5-4)

3. تحليل الانحدار:

هو عبارة عن اسلوب احصائي يقوم على اساس صياغة دالة رياضية لعملية ذات عوامل مؤثرة عدة مثل $(X_1, X_2, X_3, \ldots\ldots, X_n)$ لوصف متغير النتج من هذه العملية يروز له بالرمز (y) وكيفية التحكم بهذا المتغير وتوقع قيم غير معروفة له. شكل الدالة الرياضية تكون على الشكل التالي:

$$y = f(X_1, X_2, X_3, \ldots\ldots, X_n)$$

يمكن توضيح هذه العملية كما في الشكل التالي:



تسمى المتغيرات $(X_1, X_2, X_3, \ldots\ldots, X_n)$ بالمتغيرات المستقلة , والمتغير الناتج (y) بالمتغير التابع. تعرف هذه الدالة , بدالة الانحدار , وابسط حالة لهذه الدالة عندما يكون للعملية متغير مستقل واحد فقط يرتبط مع المتغير التابع بعلاقة خط مستقيم وكما يلي:

One. معادلة خط انحدار (y) عن (X) تعرف بالعلاقة التالية :

$$y = b + r\,X$$

ولايجاد اي قيمة مقدرة جديدة الى **(y)** نعوض بقيمة معلومة للمتغير المستقل **(X)** .

$$b = \bar{y} - r\,\bar{X}$$

نحتاج في هذه العلاقة الى ايجاد قيمة كل من (a) و (b) :

$$a = \frac{\sigma y}{\sigma X}\ \ x\ \mathrm{r}$$

$$\mathbf{a} = \frac{\sum X\,y\ - n\,(\bar{X}\,x\,\bar{y})}{\sum(X)^2\ -\ n\,(\bar{X})^2}$$

حيث ان :

$\sigma y = \ \ y$ قيمة الانحراف المعياري للمتغير

$\sigma x = \ \ X$ قيمة الانحراف المعياري للمتغير

$r =$ معامل الارتباط

$(\bar{X})\ = X$ قيمة الوسط الحسابي للمتغير

$(X) = X$ قيمة المتغير

$(y) = y$ قيمة المتغير

لايجاد قيمة b نستخدم العلاقات التالية :

$$b =\ (\bar{y})\ - r\ \ (\bar{X})$$

حيث ان :

$(\bar{X}) = X$ قيمة الوسط الحسابي للمتغير

$(\bar{y}) = y$ قيمة الوسط الحسابي للمتغير

$a = X$ معامل المتغير

تستخدم هذه العلاقات الرياضية الاحصائية لغرض توقع او التنبؤ بقيمة المتغير (y) اذا تم معرفة قيمة المتغير (X) .

Two. معادلة خط انحدار (X) عن (y) تعرف بالعلاقة التالية :

$$X = c\,x\,(y) + d$$

نحتاج في هذه العلاقة الى ايجاد قيمة كل من (c) و (d) :

$$c = \frac{\sigma y}{\sigma x} \;\; x \;\; \mathrm{r}$$

$$\mathrm{c} = \frac{\sum X\,y - n\,(\bar{X}\,x\,\bar{y})}{\sum (X)^2 - n\,(\bar{X})^2}$$

لايجاد قيمة (d) :

$$d = (\bar{y}) = c\,x\,(\bar{x})$$

تستخدم هذه العلاقة الرياضية الاحصائية للتنبؤ بقيمة (x) اذا علمت قيمة (y) , أو انحدار (x) على (y) .

ولايجاد اي قيمة مقدرة جديدة للمتغير التابع (y_n), نعوض بقيمة معلومة للمتغير المستقل (X_n), في المعادلة.

ملاحظة : (الخطأ في التنبؤ = القيمة الحقيقية – القيمة المتنبأ بها)

مثال (1) : اذا كان معامل الارتباط بين نتائج الطلبة في امتحان الاحصاء (X) وامتحان الرياضيات (y) يساوي $(r = 0.7)$, حيث ان :

الوسط الحسابي لنتائج الطلبة في الاحصاء $= \bar{X} = 60$

الوسط الحسابي لنتائج الطلبة في الاحصاء $= \bar{y} = 55$

الانحراف المعياري لنتائج الطلبة في الاحصاء $\sigma X = 7$

الانحراف المعياري لنتائج الطلبة في الاحصاء $\sigma y = 11$

المطلوب :

1. ايجاد معادلة خط انحدار نتائج الرياضيات (y) على نتائج الاحصاء (X) .

2. أوجد نتيجة احد الطلبة المتوقعة في امتحان الرياضيات , اذا كانت نتيجة الطالب في امتحان الاحصاء تساوي $(\mathbf{65})$.
3. أوجد قيمة $(\boldsymbol{X})$ المتوقعة في امتحان الاحصاء اذا علمت ان قيمة $(\boldsymbol{y})$ في امتحان الرياضيات لاحد الطلبة تساوي $(\mathbf{60})$.

الحل :

1. معادلة خط انحدار $(\boldsymbol{y})$ على $(\boldsymbol{X})$ تساوي : $\boldsymbol{y = a\,X + b}$
   a. نوجد قيمة (a) : $\boldsymbol{a = \frac{\sigma y}{\sigma X}\ x\ r = \frac{11}{7}\ x\ 0.7 = 1.1}$
   b. نوجد قيمة (b) : $\boldsymbol{b = \bar{y} - a\ x\ \bar{X} = 55 - (1.1\,x\,60) = 55 - 66 = -11}$
   c. نعوض القيم اعلاه في معادلة خط انحدار $(\boldsymbol{y})$ على $(\boldsymbol{X})$:

$$\boldsymbol{y = a\,X + b = (1.1\ x\ X) + (-11)}$$

2. لايجاد نتيجة الطالب المتوقعة في امتحان الرياضيات , اذا كانت نتيجته في امتحان الاحصاء هي (60), اي عندما تكون (X=60) , كم تكون قيمة y) .

$$\boldsymbol{y = a\,X + b = (1.1\ x\ X) + (-11) = (1.1\ \text{x}\ 65) - 11 = 60.5}$$

وهي النتيجة المتوقعة في امتحان الرياضيات

3. لايجاد قيمة (X) المتوقعة , اذا كانت (y = 60) , بهذه الحالة يجب ان نوجد معادلة انحدار (X) على (y) :

$$X = c\ x\ y + d$$

   a. يجب ان نوجد قيمة (c) :

$$c = \frac{\sigma y}{\sigma X}\ x\ r = \frac{7}{11}\ x\ 0.7 = 0.448$$

   b. نوجد قيمة (d) :

$$d = \bar{X} - c\,x\,\bar{y} = 60 - (0.448)\,x\,55 = 35.36$$

   c. نعوض في معادلة خط انحدار (X) على (y) :

$$X = c\ x\ y + d = (0.448\ x\ 60) + 35.36 = 62.24$$

مثال (2) : اذا كانت معادلة خط الانحدار للمتغيرين X , Y هي : (X على y) $\text{X} = 2\ x\ \text{y} + 90 =$
وحيث ان : ($6\ x\ \text{X} = 15$) , ($6\ x\ \text{y} = 6$) . أوجد قيمة معامل الارتباط (r) .

الحل :

بما ان معادلة خط انحدار X على y هي : $(X = c\ x\ y + d)$, وعندما نعوض في المعادلة هذه نحصل على : $(X = 2\ x\ y + 90)$, وهذا يعني ان قيمة (d = 90) و قيمة (c = 2) .

معادلة معامل الارتباط (r) هي:

$$c = \frac{\sigma X}{\sigma Y}\ x\ r\ ,\quad 2 = \frac{15}{6}\ x\ r\ ,\quad \therefore r = \frac{12}{15} = 0.8$$

مثال (3) : اذا علمت ان :

$$\sum Xy = 198\ ,\sum = 196\ ,\sum (X)^2 = 360\ ,\sum X = 93\ ,\quad \sum y = 62$$

n =31

أوجد معادلة خط انحدار (X) على (y) .

الحل:

a. نوجد قيمة (c) : $c = \frac{\sum Xy - n\ x\ \bar{X}\ x\ \bar{y}}{\sum (y)^2 - n\ x\ (\bar{y})^2}$

نستخرج الوسط الحسابي للمتغيرات X و y وكما يلي :

$$\bar{y} = \frac{\sum y}{n} = \frac{62}{31} = 2$$

$$\bar{X} = \frac{\sum X}{n} = \frac{93}{31} = 3$$

$$c = \frac{198 - (31\ x\ 3\ x\ 2)}{196 - 31\ x\ (2)^2} = 0.16$$

b. نوجد قيمة (d) :

$$d = \bar{y} - c\ x\ \bar{X} = 2 - (0.1\ x\ 3) = 1.52$$

c. نوجد معادلة خط انحدار X على y وهي :

$$X = c\ x\ y + d = 0.16\ x\ y + 1.52$$

<u>مثال(3):</u>

لدراسة علاقة الاستهلاك المحلي (y) بالانتاج لمادة وقود البنزين (بالمليون برميل) خلال عدة سنوات , تم الحصول على عشرة قراءات كما مبينة في الجدول التالي:

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

أوجد معادلة خط الانحدار البسيط , ثم اوجد قيمة توقع الاستهلاك المحلي عندما يصل الانتاج الى (16000000) برميل من وقود البنزين؟

الحـل:

1. يتم عمل جدول لغرض استكمال البيانات المطلوبة وكما يلي:

x	y	xy	x^2
10	6	60	100
13	8	104	169
15	9	135	225
14	8	112	196
9	7	63	81
7	6	42	49
6	5	30	36
6	6	36	36
5	5	25	25
5	5	25	25
$\sum x = 90$	$\sum x = 65$	$\sum x = 632$	$\sum x = 942$

2. نوجد الوسط الحسابي والانحراف المعياري للمتغيرين (x , y):

$$\bar{X} = \frac{\sum X}{n} = \frac{90}{10} = 9 \quad , \quad \bar{X} = \frac{\sum y}{n} = \frac{65}{10} = 6.5$$

$$S_x^2 = \frac{\sum X^2}{n} - \bar{X}^2 = \frac{942}{10} - 9 = 94.2 - 81 = 13.2$$

$$\text{معامل الارتباط} = r = \frac{\frac{\sum Xy}{n} - \bar{X}\bar{y}}{S_x^2} = \frac{\frac{632}{10} - (9)(6.5)}{13.2} = \frac{63.2 - 58.5}{13.2} = 0.36$$

$$b = \bar{y} - r\,\bar{X} = 6.5 - (0.36)\,(9) = 6.5 - 3.24 = 3.26$$

∴ معادلة خط الانحدار البسيط في هذه الحالة هي: $y = 3.26 + 0.36\,X$, ولتوقع قيمة الاستهلاك المحلي من مادة البنزين عندما يصل الانتاج الى (16000000) برميل , يجب تحويل وحدة هذه القيمة من برميل الى مليون برميل وذلك بالقسمة على مليون, اي ان القيمة المستخدمة في توقع الاستهلاك هي (X=16) ,

وبالتعويض في المعادلة $\boldsymbol{y = b + r\,X}$ نجد ان :

$$\boldsymbol{y = b + r\,X} = 3.26 + (0.36)(16) = 9.02$$

وهذا يعني ان الاستهلاك المحلي قد يصل الى (9.02) مليون برميل , اي ما يعادل (9020000) برميل خلال السنة.

<u>مثال(4):</u>

لدراسة العلاقة بين الدخل (X) والاستهلاك (y) بآلاف الدنانير , كانت لدينا النتائج التالية:

$$\sum X = 120\ ,\sum y = 100\ ,\sum Xy = 516\ ,\sum X^2 = 720\ ,\sum y^2 = 410\ ,n = 40$$

<u>المطلوب:</u>

1. احسب معامل الارتباط الخطي بين الظاهرتين, ثم بين ما هو نوع الارتباط ؟ وما مدى قوته؟
2. اوجد معادلة (خط) انحدار الاستهلاك على الدخل؟
3. تقدير الاستهلاك عندما يصل الدخل الى (10000) دينار

<u>الحـــل:</u>

.1 $\bar{X} = \frac{\sum X}{n} = \frac{120}{40} = 3\ \ ,\ \ S_x = \sqrt{\frac{\sum X^2}{n} - \bar{X}^2} = \sqrt{\frac{720}{40} - (3)^2} = \sqrt{18-9} = \sqrt{9} = 3$

$$\bar{y} = \frac{\sum y}{n} = \frac{100}{40} = 2.5\ \ ,S_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{410}{40} - (2.5)^2} = \sqrt{10.25 - 6.25}$$
$$= \sqrt{4} = 2$$

.2 $$\boldsymbol{r = \frac{\frac{\sum Xy}{n} - \bar{X}\bar{y}}{S_x S_y} = \frac{\frac{516}{40} - (\bar{3})(\overline{2.5})}{\sqrt{9}\sqrt{4}} = \frac{12.9-7.5}{6} = \frac{5.4}{6} = 0.9}$$

من الملاحظ ان الارتباط طردي قوي بين الدخل والاستهلاك.

$$r = \frac{\frac{\sum Xy}{n} - \bar{X}\bar{y}}{S_x^2} = \frac{\frac{516}{40} - (3)(2.5)}{9} = \frac{5.4}{9} = 0.6$$
$$y = \bar{y} - r\bar{X} = 2.5 - (0.6)(3) = 2.5 - 1.8 = 0.7$$

3. نلاحظ ان وحدة القياس هي الاف الدنانير لذلك فان قيمة الدخل (10000) دينار ستحول الى (10) الاف دينار وعليه فان (X=10) , اي ان:

$$y = b + r\,X = 0.7 + (0.6)(10) = 0.7 + 6 = 6.7$$

يعني ذلك ان قيمة الاستهلاك المقدرة تساوي 6700 دينار.

مثال(5):

البيانات المدرجة في الجدول التالي تمثل اعمار عينة من الازواج وزوجاتهم بالسنوات:

عمر الزوج (y)	50	60	24	30	25	35	44	56	37	30
عمر الزوجة (X)	40	37	20	25	19	25	25	42	30	20

المطلوب:

1. احسب معامل الارتباط الخطي بين الظاهرتين بين اعمار الزوج والزوجة.
2. احسب معامل ارتباط الرتب بين اعمار الزوج والزوجة.
3. حدد نوع وقوة الارتباط من خلال معاملي الارتباط.

الحـل:

1. نعمل على اكمال البيانات في الجدول التكراري التالي:

X	y	Xy	X^2	y^2
40	50	2000	1600	2500
37	60	2220	1369	3600

20	24	480	400	576
25	30	750	625	900
19	25	475	361	625
25	35	875	625	1225
25	44	1100	625	1936
42	56	2352	1764	3136
30	37	1110	900	1369
20	30	600	400	900
$\sum X = 283$	$\sum X = 283$	$\sum X = 283$	$\sum X = 283$	$\sum X = 283$

$$\bar{X} = \frac{\sum X}{n} = \frac{283}{10} = 28.3 \quad , \quad S_x = \sqrt{\frac{\sum X^2}{n} - \bar{X}^2} = \sqrt{\frac{8669}{10} - (28.3)^2}$$

$$= \sqrt{866.9 - 800.89} = \sqrt{66.01}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{391}{10} = 39.1 \quad , S_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{16767}{10} - (39.1)^2}$$

$$= \sqrt{1676.7 - 1528.81} = \sqrt{149.89}$$

$$r = \frac{\frac{\sum Xy}{n} - \bar{X}\bar{y}}{S_x S_y} = \frac{\frac{11962}{10} - (28.3)(39.1)}{\sqrt{66.01}\sqrt{149.89}} = \frac{1196.2 - 1106.53}{99.47} = 0.90$$

يتم تكوين جدول الرتب وكما يلي:

X	y	رتب X	رتب y	d	d^2
50	40	8	9	-1	1
60	37	10	8	2	4
24	20	1	2.5	-1.5	2.25

30	25	3.5	5	-1.5	2.25
25	19	2	1	1	1
35	25	5	5	0	0
44	25	7	5	2	4
56	42	9	10	-1	1
37	30	6	7	-1	1
30	20	3.5	2.5	1	1
				$\sum d = 0.0$	$\sum \boldsymbol{d^2} = \mathbf{17.5}$

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\ x\ 17.5}{10(100-1)} = 1 - \frac{105}{990} = 1 - 0.11 = 0.89$$

2. نلاحظ ان كلا المعاملين قيمتهما تدل على ان الارتباط طردي قوي بين اعمار عينة الازواج وزوجاتهم.

الباب الرابع

نظرية المعاينة

الفصل السابع عشر

نظرية المعاينة Sampling Theory

المقدمة:

تمثل البيانات المادة الاساسية في اية دراسة احصائية , وعلى هذا الاساس تعتبر مرحلة جمعها من اهم مراحل البحث العلمي عند تطبيق الاسلوب الاحصائي. ترتبط دقة ومصداقية البيانات المستخدمة بدقة وفعالية هذه المرحلة والتي تعتمد عليها كل مراحل التحليل الاحصائي اللاحقة , مما يؤثر على اهمية النتائج المستخرجة فضلا عن جودة القرارات المتخذة على اساس هذه النتائج , آخذين بنظر الاعتبار الظروف الميحطة بعملية البحث , فكلما كانت عملية جمع البيانات دقيقة كلما زادت ثقة الباحث في الاعتماد عليها وعلى النتائج المستحصلة منها.

تهتم نظرية المعاينة بدراسة العلاقة بين المجتمع والعينات المسحوبة منه فيما يسمى بالاستدلال الإحصائي ، Statistical Inference. هناك عدة طرق لأخذ العينات من المجتمع لاستخدامها في الاستدلال الإحصائي ومن أشهر هذه الطرق هي العينة العشوائية وهي العينة التي تكون لكل مفرده من مفردات المجتمع نفس فرصة الاختيار في العينة. فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطه أو تباينة أو غير ذلك. أو أعطاء عينه من المرضى بارتفاع الضغط، مثلاً دواء معين ثم قياس ضغطهم قبل وبعد تناولهم لهذا الدواء لمعرفة ما إذا كان هذا الدواء مفيد في خفض الضغط أم لا.

أي مجموعات من المفردات تشترك في صفه أو صفات وتكون موضوع دراسة أو بحث فإن هذه المجموعة يطلق عليها إحصائيا مجتمع الدراسة (أو اختصاراً المجتمع Population). والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينه من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينه خلال فترة زمنية محدده...الخ.

والمجتمع قد يكون محدوداً إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة ، وقد يكون المجتمع غير محدود (لانهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار وعند دراسة صفة ما أو صفات معينه لمجتمع ما ، فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد أسلوبين:

أولاً: أسلوب الحصر الشامل (census) وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد

المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني: أسلوب المعاينة (Sampling method) وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة أو بأخرى ويطلق عليه عينه (Sample) ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أي أن إسلوب المعينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، ونجاح هذا الإسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

توزيعات المعاينة:

نفرض ان لدينا مجتمع وان مفرداته تتبع توزيعا احتماليا معيناً, ونريد اختيار عينة عشوائية حجمها (n) مفردة من هذا المجتمع وكما يلي:

نفترض اننا اخترنا عينة عشوائية اولى حجمها (n) مفردة او نموذج من هذا المجتمع, ويم حساب وسطها الحسابي فكان يساوي $(\bar{X}_1)$, ثم تم اختيار عينة عشوائية ثانية لها نفس الحجم (n) مفردة او نموذج, وتم حساب الوسط الحسابي لها وكان يساوي $(\bar{X}_2)$, ثم تم اختيار عينة عشوائية ثالثة لها نفس الحجم (n) مفردة او نموذج, وتم حساب الوسط الحسابي لها فكان يساوي $(\bar{X}_3)$, وهكذا يتم تكرار هذه العملية لجميع العينات التي يمكن سحبها من نفس المجتمع. بهذه الحالة سوف يتوفر لدينا عدد كبير من القيم للوسط الحسابي للعينات , وبالطبع لا يمكن ان نتوقع انها جميعا متساوية في القيمة, وهذه القيم سوف تكون مجتمعاً آخر عدد مفرداته اكبر بكثير من مفردات المجتمع الاصلي, وبناء على ذلك يمكن النظر الى هذا المقياس الذي هو (الوسط الحسابي) على انه نتغير عشوائي ياخذ قيماً مختلفة, ويتبع توزيعاً احتمالياً معيناً قد يختلف عن توزيع المجتمع الاصلي . هذا المجتمع الجديد الذي هو مجتمع المتوسطات الحسابية يسمى بهذه الحالة (توزيع المعاينة للوسط الحسابي), وبهذا يمكن ان نعرف توزيع المعاينة لمقياس ما على انه التوزيع الاحتمالي لمجتمع هذا المقياس.

مجتمع المتوسطات الحسابية للعينات:

يعتمد شكل ونوع التوزيع الاحتمالي لمجتمع المتوسطات الحسابية للعينات العشوائية على توزيع المجتمع الاصلي الذي اختيرت منه هذه العينات العشوائية, والنظرية الآتية تعطي التوزيع الاحتمالي لمجتمع المتوسطات الحسابية للعينات العشوائية الكبيرة.

نظرية النهاية المركزية:

اذا كان لدينا مجتمع (غير محدود) مفرداته (X) , يتبع توزيعاً احتمالياً متوسطه (μ) واحرافه المعياري (σ). تم سحب عينات عشوائية من هذا المجتمع حجم كل منهما يساوي (n) , وكانت (n) كبيرة الحجم اي $(n \geq 30)$. فان الوسط الحسابي لهذه العينات $(\bar{X})$ يتبع توزيعا طبيعياً له الخصائص التالية:

$$\mu(\bar{X}) = \mu \quad , \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

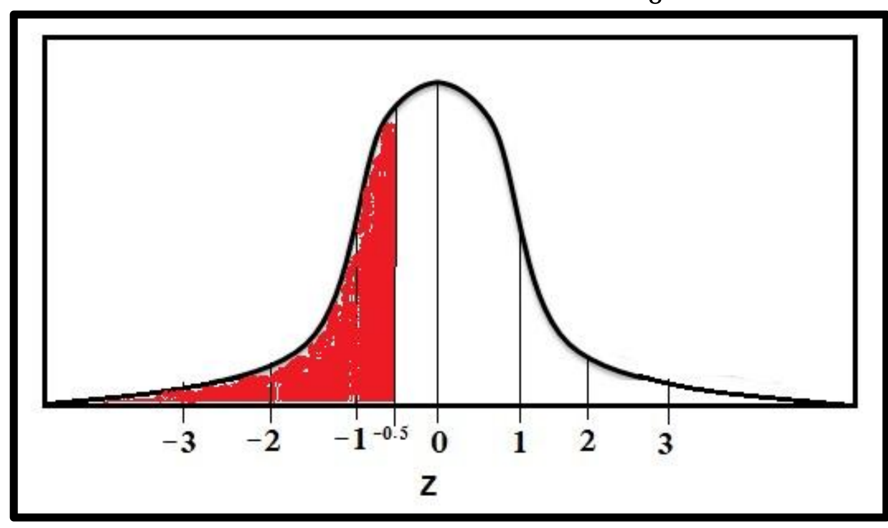
مثال(1):

اذا كان بدل السكن المعطى للموظف باحدى الشركات المساهمة يتبع توزيع طبيعي متوسطه $\mu = 170$ دولار , وانحرافه المعياري $\sigma = 8$ دولار:

1. اذا تم اختيار موظف عشوائي , فما هو احتمال ان يقل بدل سكنه عن 166 دولار؟
2. سحبت عينة من (64) موظف , فما هو احتمال ان يكون متوسط بدل سكنهم اكبر من 172 دولار ؟

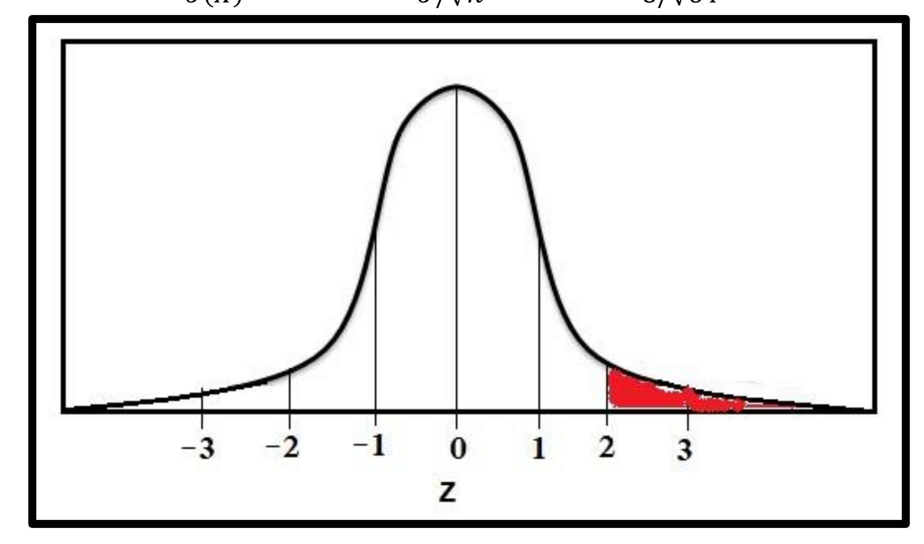
الحـــــل:

.1 $X < 166 \rightarrow Z < \frac{166-170}{8} \rightarrow Z < -0.5$



$$P(X < 166) = P(Z < -0.5) = 0.5 - 0.1915 = 0.3085$$

$$\bar{X} > 172 \rightarrow Z > \frac{172-\mu(\bar{X})}{\sigma(\bar{X})} \rightarrow Z > \frac{172-\mu}{\sigma/\sqrt{n}} \rightarrow Z > \frac{172-170}{8/\sqrt{64}} \rightarrow Z > 2$$ .2



$$P(\bar{X} > 172) = P(Z > 2) = 0.5 - 0.4772 = 0.0228$$

الفصل الثامن عشر

البيانات والمتغيرات Data and Variables

البيانات الاحصائية The Statistics Data:

تلعب البيانات الاحصائية دورا مهماً في حياتنا المعاصرة , فهي الادة الاساسية التي تعتند عليها جميع البحوث والدراسات ووسيلة هامة للتعبير الكمي والنـءعي عن الظواهر والمشكلات المطروحة من جهة وكذلك لترشيد عملية اتخاذ القرارات على جميع المستويات والتنبؤ بسلوك هذه الظواهر وتطورها المستقبلي من جهة اخرى. يمكن تعريف البيانات الاحصائية بانها مجموعة من المشاهدات او الملاحظات التي تؤخذ اثناء دراسة معينة , او هي مجموعة من الارقام لو الحقائق الرقمية التي تحتاج الى معالجة وتنظيم أو اعادة تنظيم لكي تتحول الى معلومات , ويمكن التعبير عنها بانها المادة الخام , يتم تحويلها الى معلومات قابلة للاستعمال بالاعتماد على مجموعة من الوسائل والاساليب الاحصائية الوصفية او الكمية. بصورة عامة , تمثل البيانات الاحصائية في البحث العلمي كل ما يحصل عليه الباحث من حقائق تخص الظاهرة المدروسة , وعلى هذا الاساس فهي المادة الريسية في اي بحث احصائي حيث ترتبط دقة البحث والتحليل في مدى توفر ودقة البيانات الاحصائية .

تاخذ البيانات الاحصائية في الواقع العملي صيغاً واشكالاً مختلفة مثل: الارقام , الاشكال , المؤشرات الكمية , الرسومات البيانية , او تكون مزيج من هذه العناصر . يمكن تحديد خصائص البيانات الاحصائية بانقاط التالية:

a. هي عبارة عن مجموعة من القيم التي يمكن الحصول عليها باساليب مختلفة (المجتمع او العينة) من مصادر متنوعة وباستخدام طرق متنوعة وادوات مختلفة مثل: المقابلة الشخصية , الاستقصاء , الملاحظة والاستبيان .

b. تعبر عن قيم فعلية يتم الحصول عليها من التجارب او الدراسات المختلفة , وقد تكون بيانات كمية قابلة للوصف والقياس والتحليل الاحصائي , وقد تكون بيانات نوعية لا يمكن قياسها الا بعد تحويلها الى قيم كمية او عددية . ويتم عرض البيانات بشكل يسهل استغلالها وتداولها كالجداول والاشكال البيانية .

c. تمثل البيانات في البحوث العلمية حقائق , والحقيقة تتصف بقدر كبير من الثبات. يجب التفرقة بين البيانات والمعلومات في ميدان البحث العلمي , فالمعلومات هي حقيقة تم اثباتها , في عبارة عن النتائج النهائية للبيانات بعد معالجتها باستخدام الاساليب الاحصائية ووسائل الاتصال والتكنولوجيا المختلفة , لكي تبنى على اساسها نظريات مسلم بها , ذلك بان البيانات الخام ليس لها معنى ولا يمكن استغلالها ما لم تعالج .

اهمية البيانات الاحصائية :

تُعد البيانات الاحصائية من اهم الموارد التي تعتمد عليها الحياة المعاصرة في مجالاتها كافة سواء على مستوى الافراد او على مستوى المؤسسات مهما كان نوعها , ويمكن تلخيص هذه الاهمية بالنقاط التالية :

1. تعطي البيانات الاحصائية صورة موضوعية في لحظة زمنية معينة عن المجتمع بمميزاته وخصائصه المختلفة .
2. البيانات الاحصائية ضرورية ومهمة لاتخاذ القرارات على جميع الاصعدة سواء بالنسبة للافراد في حياتهم اليومية او بانسبة للمؤسسات بمختلف انواعها .
3. مراقبة وتقييم مدى تطور المجتمع واظهار الفروق بين المناطق والمحافظات والاقاليم والفئات المختلفة في الوطن الواحد , وبالتالي فهي ضرورية للمجتمع المدني .
4. تعزيز مبادئ البحث العلمي والاكاديمي وتطويره والمساهمة ليس فقط في تقييم ومراقبة التقدم بل في انجازه .
5. تحقيق الميزة التنافسية بين المؤسسات والمتمثلة في قدرتها على صياغة وتطبيق الاستراتيجيات التي تجعلها في مركز افضل بالنسبة للمؤسسات الاخرى العاملة في النشاط نفسه عن طريق الاستغلال الافضل للقدرات , الكفاءات , الامكانات والموارد الفنية , المادية , المالية , التنظيمية وبالاخص المعلوماتية التي تتمتع بها المؤسسة والتي تمكنها من تصميم وتطبيق رؤيتها واستراتيجيتها التنافسية .
6. تعد البيانات الاحصائية الجيدة مدخلا عقلانيا لادارة وتوفير الخدمات الاساسية ومتابعة الاثار المترتبة على السياسات التنموية المختلفة .
7. ترتبط شفافية ومسؤولية ومتابعة صناعة القرار بجودة وقوة البيانات والمعلومات الاحصائية , والتي تعتمد على طريقة واسلوب جمعها وذلك وفقا لقواعد التطبيق الجيد والمعايير المتفق عليها لضمان تسيير موجه وفقا للنتائج المراد الوصول اليها.
8. تمد رسم السياسات والخطط واتخاذ القرارات سلبا او ايجابا على حجم وجودة المعلومات المستخدمة , وبالتالي اي نقص او خلل او سوء استعمال يؤدي الى تحمل تكاليف مالية وبشرية عالية دون الوصول الى تحقيق الاهداف .

انواع البيانات الاحصائية :

يرتبط نوع البيانات بنوع المتغير المدروس , كما تختلف طبيعة البيانات الاحصائية حسب الهدف من استخدامها. تعبر البيانات عن قيم لمتغير او اكثر من المتغيرات الاحصائية , ويمثل المتغير الاحصائي خاصية مشتركة او اكثر تميز عناصر المجتمع المدروس , التي تختلف باختلاف الظاهرة المدروسة , فضلا عن اختلاف الزمان والمكان وغيرها من العوامل التي تفيد في تحديد خصائص المجتمع المدروس الذي سيتم تحليله لاحقاً مثل دخل الفرد, الحالة الاجتماعية, الصول , الوزن, تركيز عنصر معين,.....الخ. ان كل خاصية او صفة من هذه الصفات تعتبر متغيرا احصائياً في حد ذاتها , ويرتبط نوع البيانات بنوع المتغير المدروس , وبالتالي فان انواع المتغيرات الاحصائية تمثل بدورها نوع البيانات التي يمثلها هذا المتغير.

تنقسم البيانات او المتغيرات الاحصائية الى نوعين رئيسيين وفقا للقيم التي ياخذها هذا المتغير , لذا التصنيف اهمية كبيرة حيث انه يحدد طبيعة التحليلات الاحصائية الملائمة لهذه المتغيرات وهما :

1. متغيرات رقمية (كمية) Quantitative data :

المتغيرات الكمية هي الظاهرة أو الصفة التي يمكن قياسها كمياً, بمعنى ان هذا النوع من المتغيرات يعكس كميا مدى توافر خاصية معينة . فالقيم الممكنة للمتغير الكمي قابلة للقياس كميا بارقام عددية لها خصائص حسابية وباستخدام وحدات قياس محددة , تمكننا من المقارنة الدقيقة بين قيمتين مختلفتين , فهي عبارة عن اعداد حقيقية (موجبة, سالبة او معدومة) تتعلق بكميات كالوزن , الطول, الحجم , الدخل,....الخ. تسمى البيانات التي تعبر عن مثل هذه الخصائص او عن هذا النوع من المتغيرات بالبيانات الكمية , هناك نوعين من المتغيرات الكمية هي:

a. متغيرات منفصلة او متقطعة Discrete: وهي المتغيرات المتقطعة المحدودة التي تاخذ قيماً (اعداد) او وحدات كاملة والتي يمكن عدها , مثل عدد افراد الاسرة , التي ممكن ان تكون 2 او 3 او 4 او 5 , اي هذه الاعداد صحيحة لا يوجد فيها كسور عشرية , او الكتب في المكتبة التي لا تقبل كسور عشرية بينهما, وهكذا تسمى بيانات متقطعة او منفصلة ... الخ.
b. المتغيرات المتصلة Continuous: وهي المتغيرات التي يمكن قياسها , والتي تاخذ اي قيمة في مجال معين ولا يوجد اي انفصال او انقطاع بين القيم وتاخذ قيما غير محدودة وغير معدودة, ولذلك تسمى قيم متصلة او مستمرة, مثل قياس اطوال الطلبة في المرحلة الرابعة, قياس المسافات بين المدن,...الخ وهي التي تقبل وجود كسور عشرية بين الاعداد العشرية .

2. متغيرات نوعية (وصفية) Qualitative Data:

المتغيرات النوعية , يمكن العبير عنها بانها المتغيرات او الظواهر التي لا يمكن قياسها باستخدام وحدات معينة مثل , لون العين, المستوى التعليمي, الحالة الاجتماعية ... فهي تعبر عن الحالات , الاراء, السلوك, الخصائص, صفات الاشخاص او الاشياء , وبالتالي فهي عبارة عن قيم ذات طابع نوعي (تتعلق بالنوعية), سواء اخذت قيم المتغير النوعي ترتيبا معينا او لم تاخذ , فليس لها دلالة حقيقية مثل المتغير الكمي لانه لا يوجد اي مقياس مرجعي . تسمى البيانات التي تعبر عن هذا النوع من المتغيرات (بالبيانات النوعية). يقاس هذا النوع من المتغيرات بمقياسين هما : المقياس الاسمي وتسمى المتغيرات في هذه الحالة بالمتغيرات الاسمية مثل ذكر , انثى , او لون العين , او نوع الجنسية , ...الخ. , والمقياس الترتيبي , وتسمى في هذه الحالة بالمتغيرات النوعية الترتيبية , وهذا وفقا لقابلية قيم هذا المتغير للترتيب من عدمه , مثل مثل صغير , وسط , كبير ....الخ.

انواع المتغيرات Kinds of Variables :

المتغيرات اما احصائية او عشوائية , فالمتغير الاحصائي يمثل القيم التي تاخذها ظاهرة ما , في حين ان المتغير العشوائي هو عبارة عن ظاهرة نوعية او كمية لا يمكن التنبؤ بها بشكل مسبق وتقترن بقيم احتمالية .

ملاحظة : البيانات تصبح متغيرات عند اجراء العمليات الاحصائية عليها , حيث ان المتغيرات لها نفس مفهوم البيانات , والذي ينطبق على البيانات ينطبق على المتغيرات .

a. المتغيرات الاسمية Nominal Variables: هذه المتغيرات لها عدد من الفئات المحددة من دون اي معنى كمي لهذه الفئات , مثال على ذلك يمكن تصنيف افراد المجتمع الى عدد معين من الفئات من دون ان تكون هناك افضلية لاحدى هذه الفئات على الاخرى , مثل متغير لون العين , وفي معظم الاحيان يتم اعطاء ارقام تدل على هذه الفئات , مثلا نرمز الى لون العين السوداء بالرقم (1) والى لون العين الازرق بالرقم (2) وهكذا , ولا تدخل هذه الارقام في العمليات الحسابية .

b. المتغيرات الترتيبية Ordinal Variables: وهي المتغيرات التي تمتلك عدد محدد من الفئات يمكن ترتيبها تصاعديا او تنازليا , ولكن لا يمكن تحديد الفروق بدقة بين الافراد المختلفة , مثال على ذلك كبير , صغير , وسط .

c. المتغيرات الفئوية Interval Variables: هي تلك المتغيرات الكمية التي يمكن اجراء العمليات الحسابية على قيمها دون ان تتاثر المسافة النسبية بين قيمها , وتكون قيمة الصفر لا تعني عدم توفر تلك الصفة مثال على ذلك لو حصل طالب على درجة الصفر في امتحان اللغة العربية فهذا لا يعني ان الطالب لا يعرف شيئا عن اللغة العربية , واذا كانت درجة الحرارة تساوس صفر فهذا لا يعني عدم وجود درجة حرارة .

d. المتغيرات النسبية Ratio Variables: وهي عبارة عن متغيرات كمية , ليس لها فئات محددة , وان قيمة الصفر في هذا النوع من المتغيرات يمثل عدم توفر الصفة (مثل المتغير الزمني , فاذا قلنا ان الزمن يساوي صفر , اي لا زمن هناك , وايضا عندما نقول ان المسافة تساوي صفر فهذا يعني ان لا مسافة موجودة) , فلذلك يكثر استخدام هذا المتغير فيزيائياً .

المجتمع Population : المجتمع الاحصائي هو عبارة عن جميع الوحجات او العناصر او المتغيرات التي تدخل في موضوع الدراسة , سواء كانت هذه الوحدات او العناصر او المتغيرات هي افراد او اشياء او قياسات ... الخ. مثال على ذلك دراسة اعمار طلاب جامعة ما , فالمجتمع الاحصائي هنا هو طلاب الجامعة في وقت الدراسة , وقد يكون المجتمع الاحصائي محدود وقد يكون غير محدود.

العينة او النموذج Sample: تعرف العينة على انها جزء صغير او نموذج من المجتمع (Pupolation) الكلي يلجأ اليه الباحث في دراسته , ويختار بطريقة مناسبة ويراعى ان تكون هذه العينة عشوائية وبدون انتقاء وممثلة لجميع خصائص المجتمع تمثيلا صادقا , بدقة وموضوعية ,وهذا يتطلب تحديد هدف الدراسة ومجتمع الدراسة. حيث ان العينة تسحب من المجتمع الاصلي لغرض دراسة صفاته وخصائصه , مثلا ان الطبيب يريد معرفة مرض ما في دم المريض فلا داعي لسحب كل دم المريض بل يتم اخذ عينة من الدم لفحصها.

مثال: دراسة عنوانها , الصعوبات التي تواجه طلبة المرحلة الثانية في كلية العلوم في مادة الاحصاء , حدد مجتمع الهدف , مجتمع العينة ؟

الحل:

مجتمع الهدف : جميع طلبة المرحلة الثانية في كلية العلوم.

مجتمع العينة : الجزء الذي تؤخذ من العينة , بمعنى الكليات التي اخذت منها العينة, مثل كلية العلوم , كلية الادارة والاقتصاد .

الفصل التاسع عشر

طرق او اسلوب جمع البيانات

يتم جمع البيانات باحد الطريقتين التاليتين:

1. الحصر او المسح الشامل : حيث يتم جمع البيانات من جميع افراد مجتمع الدراسة (Pupolation) , ويستخدم عادة هذا النوع من جمع البيانات في المجتمعات الصغيرة مثل (المدرسة , المصنع , ...) أو في حالة تباعد الازمنة بين الاحصائيات , مثل التعداد السكاني , وتمتاز نتائج هذه الطريقة بالقة العالية والوضوح وتكمن قوة وموثوقية الحصر الشامل في اعطاء الباحث صورة كاملة عن مجتمع الدراسة , ومن مساوئ هذا النوع من جمع البيانات هو التكاليف الباهظة التي يتطلبها وطول المدة الزمنية التي يستغرقها لانجاز العمل والحاجة الى عدد كبير من الباحثين.
2. العينات : يستخدم اسلوب العينات عند دراسة المجتمعات الكبيرة جداً .

انواع العينات او النماذج :

1. العينة العشوائية : وهي العينة التي يتم اختيارها من المجتمع بصورة عشوائية وبدون انتقاء بحيث تكون فرص ظهور اي من عناصر او مفردات او متغيرات المجتمع فيها متساوية او متكافئة, بمعنى آخر يكون كل نموذج فيها له نفس فرصة الاختيار من المجتمع الكلي.
2. العينة العمدية (المتعمدة) او القصدية : يتم اختيار هذه العينة بطريقة تناسب اهداف البحث بحيث تتوفر في كل عنصر او متغير من عناصر العينة شروط محددة يرى الباحث انها تساعده على الوصول الى نتائج افضل في دراسته (على سبيل المثال يتم اختيار الطلاب الاذكياء فقط في تطبيق الدراسة عليهم) . بمعنى اخر يتم اختيارها بصورة قصدية وغير عشوائية وذلك للحصول على معلومات لتكوين فكرة سريعة او لفحص الاستبيانقبل توزيع الاستمارة وتعميمها(لدراسة مدى صدق وموثوقيية الاستبيان).
3. العينة الطبقية : وهي العينة التي يتم اختيارها لتشتمل على خواص المجتمع باستخدام طريقة النسب , ويقسم عندما يكون المجتمع مقسم الى مجموعات بحث تتشابه افراد كل مجموعة بالصفات (تكون متجانسة) حيث تسمى كل مجموعة بالطبقة . مثال على ذلك اذا كان لدينا مجتمع تعليمي عدده (300) عنصر او فرد او نموذج , وكانت نسبة الذكور الى الاناث هي (2 : 3) واردنا ان نختار عينة من (50) شخصاً , فلا بد ان نختار (30) عنصر ذكر و(20) عنصر انثى . تحكمها في ذلك العلاقة التالية :

$$\text{عدد افراد عينة الطبقة} = \frac{\text{عدد افراد الطبقة}}{\text{عدد افراد المجتمع}} \; x \; \text{عدد افراد العينة الكلية}$$

مثال : يراد اختيار عينة مكونة من (20) طالب من طلبة احدى الكليات , اذا علمت ان عدد طلاب هذه الكلية (1000) طالب وهم مقسمين كما في الجدول ادناه بحسب السنة :

المرحلة	عدد الطلاب
الاولى	400

الثانية	300
الثالثة	200
الرابعة	100
المجموع	1000

بناء على ذلك , كوّن العينة المطلوبة ؟

الحل : من العلاقة اعلاه يتم استخراج عدد الطلاب (العينة المطلوبة) من كل مرحلة وكما يلي:

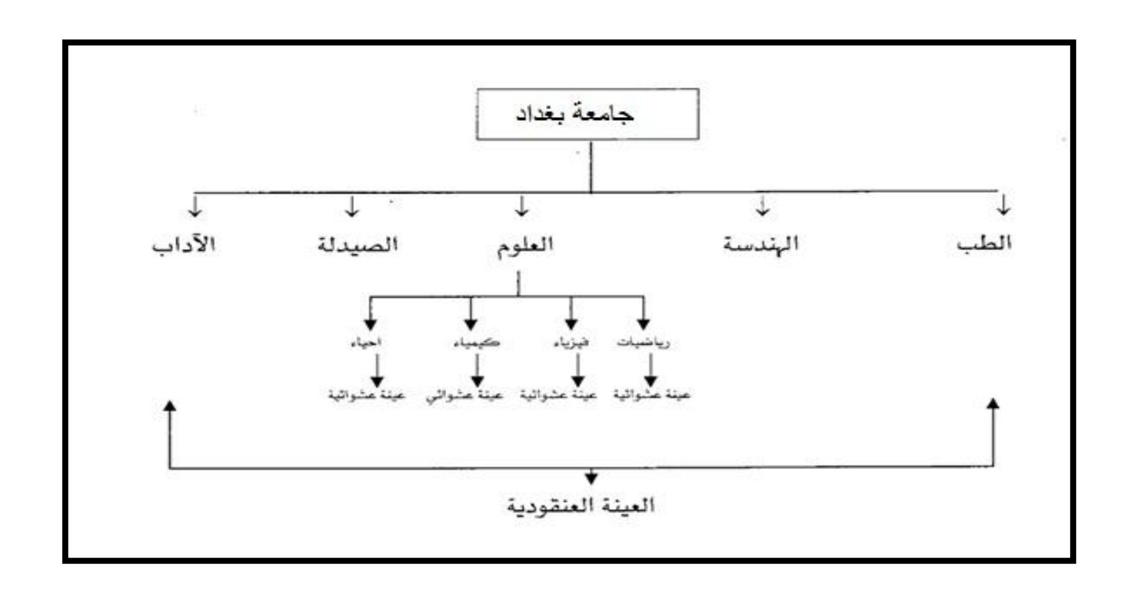
1. عدد افراد عينة الطبقة $= \frac{\text{عدد افراد الطبقة}}{\text{عدد افراد المجتمع}}\ x$ عدد افراد العينة الكلية $= \frac{400}{1000}\ x\ 20 = 8$ نختار (8) من (400) حسب العينة العشوائية البسيطة من (000) الى (399) .
2. عدد افراد عينة الطبقة $= \frac{\text{عدد افراد الطبقة}}{\text{عدد افراد المجتمع}}\ x$ عدد افراد العينة الكلية $= \frac{400}{1000}\ x\ 20 = 6$ نختار (6) من (300) حسب العينة العشوائية البسيطة من (000) الى (299) .
3. عدد افراد عينة الطبقة $= \frac{\text{عدد افراد الطبقة}}{\text{عدد افراد المجتمع}}\ x$ عدد افراد العينة الكلية $= \frac{200}{1000}\ x\ 20 = 4$ نختار (4) من (200) حسب العينة العشوائية البسيطة من (000) الى (199) .
4. عدد افراد عينة الطبقة $= \frac{\text{عدد افراد الطبقة}}{\text{عدد افراد المجتمع}}\ x$ عدد افراد العينة الكلية $= \frac{100}{1000}\ x\ 20 = 2$ نختار (2) من (100) حسب العينة العشوائية البسيطة من (000) الى (99) .

هذه العينات هي التي نطلق عليها العينة العشوائية البسيطة .

4. العينة العنقودية (متعددة المراحل): وفي هذا النوع من العينات يتم تقسيم المجتمع الى مجموعات جزئية لا يشترط تجانسها , وهذه المجموعات الجزئية تنقسم الى مجموعات جزئية اخرى وهكذا , بحيث تسمى اصغر مجموعة جزئية بالعنقود ومن ثم يتم اختيار من كل عنقود عتنة عشوائية بسيطة لتتشكل في النهاية عينة عنقودية .

مثال : دراسة فرص عمل طلاب جامعة بغداد بعد التخرج , حدد افضل عينة ؟

الحل : العينة يجب ان تكون عنقودية لان هناك طلاب جامعة ⟸ طلاب كليات ⟸ تخصصات كل كلية



5. العينة المنتظمة : ويستخدم هذا النوع من العينات عندما لا يتوفر لدينا قوائم او اية معلومات عن عدد عناصر المجتمع , ويتم بهذه الحالة اختيار افراد العينة بشكل منتظم .

مثال : اعط دراسة عن مدى رضا طلاب الجامعة عن خدمة المواصلات من والى الجامعة ؟

الحل : هنا لا نعرف عدد الطلاب الذين يستخدمون المواصلات من والى الجامعة , لذا يقف يجب ان يقف الباحث عند باب الجامعة ويختار مثلا طالب واحد من كل (50) وكما يلي:

الطالب الاول , طالب رقم 50 , طالب رقم 100, طالب رقم 150 , وهكذا حيث تكون الزيادة بين كل عينة والذي يليه ثابتة .

6. العينة المعيارية : وهي العينة الاكثر صدقا وموثوقية في تمثيل المجتمع الاحصائي.

مثال : مصنع للادوية يراد دراسة مدى فعالية دواء معين للشفاء من مرض معين ؟

الحل : يطبق الدواء على اول (10) مرضى وترصد فعاليته.

يطبق الدواء على اول (20) مرضى وترصد فعاليته.

يطبق الدواء على اول (30) مرضى وترصد فعاليته.

نستمر في التجربة الاحصائية حتى يثبت الدواء فعاليته فيعمم لعلاج المرض .

<u>مصادر البيانات الاحصائية :</u> تعتبر مرحلة جمع البيانات من اهم مراحل البحث الاحصائي , فاذا تميزت بالدقة والموثوقية انعكس ذلك على دقة التحليل الاحصائي وصحة النتائج والاستنتاجات الاحصائية . تنقسم مصادر جمع البيانات اللازمة لاي دراسة احصائية الى نوعين هي:

A. المصادر التاريخية :

وتسمى احياناً بالبيانات الثانوية, وهي المصادر التي قامت بجمعها ونشرها هيئات سواء كانت محلية او مركزية ,حكومية او غير حكومية , وطنية او دولية تتعلق بالظاهرة محل الدراسة . وهو ما يؤخذ من الكتب , الدوريات , السجلات المحفوظة مثل سجلات المواليد والوفيات , وكذلك البيانات التي تتضمنها رسائل الماجستير واطاريح الدكتوراه , مواقع الانترنيت والنشرات الاحصائية , فضلا عن البيانات التي يتم نشرها من قبل المؤسسات والمنظمات الدولية ... الخ .

تكون هذه المصادر تحت تصرف الباحث دون الحاجة الى جمع البيانات من وحدات المجتمع الاحصائي مباشرة , وبهذا تكون المصادر التاريخية للبيانات هي كل البيانات الجاهزة التي يعتمد عليها الباحث ويستخدمها بطريقة غير مباشرة ولا يعتبر مسؤولا عنها بل الجهة التي جمعتها ونشرتها هي المسؤولة عنها وتنحصر مسؤولية الباحث في هذه الحالة في تحليل هذه البيانات واستخلاص النتائج . ينقسم هذا النوع من المصادر الى قسمين :

أ. المصادر الداخلية : تمثل ادارات واقسام المنظمة محل الدراسة من خلال سجلاتها ووثائقها التي تتعلق بانشطتها المختلفة مثل (المالية , الانتاجية , المرارد البشرية , البحوث العلمية ....) وتعتبر عملية تسجيل وحفظ البيانات الاساسية الخاصة بالمؤسسة مهمة وضرورية حتى تستطيع توفير قاعدة بيانات تمكن الباحثين من الاستفادة منها في دراسة المشكلات واتخاذ القرارات المناسبة بذلك .

ب. المصادر الخارجية : هنالك العديد من المصادر الخارجية للبيانات والتي يمكن ان يعتمد عليها الباحث عند دراسة مشكلة معينة , ومن امثلة هذه المصادر , هي الوزارات ومؤسسات الدولة المختلفة , الجامعات ومراكز البحوث , الاجهزة الاحصائية وغيرها من قواعد البيانات .

B. المصادر الميدانية :

وتسمى ايضاً بالبيانات الاولية , وهي تلك المصادر التي لها علاقة مباشرة بالظاهرة المدروسة , يتم جمع البيانات بطريقة مباشرة عن طريق اتصال الباحث بالمجتمع محل الدراسة مباشرة , حيث يتم جمع البيانات ميدانياويقوم الباحث بجمع البيانات بنفسه من وحدات المجتمع او العينة محل الدراسة بعد معرفته باهداف الدراسة ونوعية المعلومات المطلوبة , تجمع هذه البيانات بعدة وسائل هي :

أ. المقابلة الشخصية : بهذه الحالة يقوم الباحث باجراء مقابلة مباشرة مع افراد المجتمع المراد دراسته مع توجيه الاسئلة الواردة في البطاقة الاحصائية لكل فرد وتسجيل اجاباته , بهذه الحالة يستطيع الباحث ان يحقق اعلى درجات الدقة في جمع البيانات . من مساوئ هذه الطريقة انها تستغرق وقتاً طويلاً وجهد وتكاليف مادية كبيرة .

ب. طريقة الاستقصاء (الاستبيان) :: وهي من اكثر طرق جمع البيانات الاولية استخداما وتعتند هذه الطريقة اساساً على تصميم مجموعة من الاسئلة التي يتوجب الاجابة عليها من طرف المستقصى منه . ويعرف الاستقصاء بانه ذلك الاسلوب المنهجي المنظم لجمع البيانات من الاطراف المستهدفة بغرض الفهم او التنبؤ في بعض الظواهر الخاصة بالمجتمع او الظاهرة المدروسة , حيث يقوم الباحث بتصميم وتهيئة استمارة تشتمل على اسئلة محددة تحقق اهداف البحث , مع مراعاة شروط كتابة الاستمارة من حيث الدقة والوضوح في عباراتها وان لا تشتمل على عبارات او اسئلة مكررة وان يركز الباحث فيها على الهدف من اجراء البحث والغاية المنشودة التي يمكن الحصول عليها من البحث وان تكون لاسئلة مرتبة من الاسهل الى الاصعب , وهكذا .

ت. طريقة الملاحظة : وهي مراقبة وتسجيل البيانات عن سلوك الظاهرة المدروسة وتستخدم هذه الطريقة في حالة استحالة الحصول على البيانات عن طريق الاستقصاء او المقابلة . تتم الملاحظة عن طريق الافراد او آليا باستخدام الاجهزة والمعدات المتخصصة بالفحص والمراقبة .

الفصل العشرون

المعلمة والاحصائية :

a. المعلمة : هو عبارة عن شئ يميز المجتمع ككل , مثل متوسط الدخل الشهري في دولة معينة , او متوسط طول الطلاب في مدرسة معينة , او نسبة الامية في مجتمع ما , ...الخ.

b. الاحصائية : وهي عبارة عن شئ يميز العينة , مثل متوسط الدخل الشهري لعينة مكونة من 100 اسرة في دولة ما , او متوسط الطول لعينة مكونة من 30 طالب في مدرسة ما , وهكذا .

المقاييس الإحصائية التي تحسب من بيانات مجتمع الدراسة بأكمله يطلق عليها معالم المجتمع (Parameters of population)، أما المقاييس الإحصائية التي تحسب من بيانات عينه مسحوبة من مجتمع الدراسة فيطلق عليها إحصاءات (Statistics) ويعتبر كل إحصاء منها بمثابة تقدير أو قيمه تقديرية لمعلمة المجتمع المناظر، فيكون المتوسط الحسابي المحسوب من بيانات العينة تقدير لمعلمة المجتمع المناظرة وهي المتوسط الحسابي المسحوب منه هذه العينة وهكذا.. ويجب ألا يغيب عن الأذهان بأن حساب قيمة المتوسط الحسابي من بيانات العينة ليس هدفاً في حد ذاته ولكن وسيلة للتعرف على المتوسط الحسابي للمجتمع موضوع الدراسة.. وهكذا بالحال بالنسبة لباقي المقاييس الإحصائية التي تحسب من العينة.

للتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز μ بينما يرمز للمتوسط الحسابي للعينة بالرمز $\bar{x}$ ، أيضاً للانحراف المعياري للمجتمع بالرمز σ بينما يرمز للانحراف المعياري للعينة بالرمز S وهكذا.

مثال : ناقش العبارة التالية : استخدام العينات هو الاسلوب الاكثر استخداما في البحوث ومفضل على اسلوب المسح الشامل .

الحل :

1. المسح الشامل يؤدي الى فساد عناصر المجتمع في بعض البحوث .
2. توفير الوقت والجهد والنفقات في اسلوب العينة .
3. المسح الشامل يحتاج الى اعداد كبيرة من الباحثين ولعدم توفرهم يضطر الباحث الى الاستعانة باشخاص قليلوا التدريب مما يزيد من نسبة الاخطاء.
4. الحاجة في بعض البحوث الى النتائج بسرعة لاتخاذ القرار .
5. تعذر الوصول الى جميع افراد المجتمع

توزيعات المعاينة Sampling Distributions :

نفرض أننا أخذنا عينه حجمه (n) من مجتمع ما ، ثم سحبنا منها بعض المقاييس الإحصائية مثل المتوسط الحسابي ، التباين ، ... فإن كل مقياس من هذه المقاييس يعتبر متغير عشوائي في ذاته يختلف من عينه إلى أخرى , هذا المتغير العشوائي يخضع لتوزيع معين , هذا التوزيع يسمى بتوزيع العينة. فمثلاً نقول أن توزيع المعاينة للمتوسط الحسابي وهو عبارة عن توزيع جميع المتوسطات الحسابية للعينات المأخوذة من نفس هذا المجتمع ذات الحجم (n)، وكذلك فإن توزيع المعاينة للتباين هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم (n) ومأخوذة من نفس المجتمع ، وهكذا ... .

توزيعات المعاينة للأوساط Sampling Distributions of Mean:

نفرض أننا سحبنا عينه حجمها (n) من مجتمع لانهائي ، القيمة المتوقعة له تساوي (μ) والانحراف المعياري هو (σ) فإن المتوسط الحسابي $(\bar{x})$ يخضع لتوزيع ما ، متوسط هذا التوزيع وانحرافه المعياري هما:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad , \quad \mu_{\bar{x}} = \mu \quad \ldots\ldots\ldots\ldots \quad (20-1)$$

وفي الحالة التي يكون فيها المجتمع الأصلي المسحوبة منه العينة مجتمع طبيعي (ويرمز له بالرمز $N\ (\mu, \sigma^2)$ فإن توزيع المعاينة للمتوسط الحسابي $(\bar{x})$ يكون في هذه الحالة توزيع طبيعي أيضاً له نفس المتوسط الأصلي (μ) ولكن انحرافه المعياري يساوي $\sigma/\sqrt{n}$ ، أي بمعنى أن:

$$X \sim N\ (\mu\ , \sigma^2) \ \rightarrow\ \bar{x} \sim N\left(\mu\ , \frac{\sigma^2}{n}\right) \quad \ldots\ldots\ldots \quad (20-2)$$

ومن ثم يكون :

$$z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0,1) \quad \ldots\ldots\ldots \quad (20-3)$$

أما إذا كان المجتمع غير طبيعي فإن $(\bar{x})$ لا تخضع للتوزيع الطبيعي ولكنها تتوزع توزيع يكون قريباً من التوزيع الطبيعي لقيم (n) الكبيرة $(n \geq 30$ حيث أن:

$$z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \quad \frac{as}{n \rightarrow \infty} \quad N(0,1) \quad \ldots\ldots\ldots \quad (20-4)$$

وتعتبر النتيجة السابقة الهامة جداً في الإحصاء وخاصة في التطبيقات العلمية وتسمى نظرية النهاية المركزية Central Limit Theorem والتي تنص على أنه في حالة العينات الكبيرة الحجم فإن المتوسط الحسابي$(\bar{x})$ يخضع للتوزيع الطبيعي بالمعاملات (μ) و$\left(\frac{\sigma^2}{n}\right)$ ، حيث أن (μ) و (σ^2) هما متوسط وتباين

المجتمع الأصلي بغض النظر عن شكل توزيع المجتمع الأصلي. ومن ثم فإنه لقي(n) الكبيرة تتحقق العلاقة (3−20) بصرف النظر عن توزيع المجتمع الأصلي.

كذلك فإنه إذا كان (x_1) هو المتوسط الحسابي لعينه عشوائية مسحوبة من مجتمع لانهائي متوسطه هو (μ_1) وانحرافه المعياري هو (σ_1)، وكان (x_2) هو المتوسط الحسابي لعينة عشوائية مسحوبة من مجتمع لانهائي آخر متوسطه (μ_2) وانحرافه المعياري (σ_1) وكانت العينتين مستقلتين فإن المجموع الجبري لمتوسط العينتين يخضع لتوزيع المعاينة بالمعاملات:

$$\mu_{(x_1+x_2)=}(\mu_1) \pm (\mu_2) \quad and \;\; \sigma^2{}_{(x_1+x_2)} = \frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2} \;\; \ldots\ldots\ldots. \quad (20-5)$$

حيث ان : $(n_2 \;,\; n_1)$ هما حجم العينة الأولى والثانية.

وإذا كان المجتمعين الأصليين طبيعيين فإن $(\bar{x}_1 \pm \bar{x}_2)$ يخضع لتوزيع طبيعي أيضاً بالوحدات المعطاة في العلاقة (5−20) وعليه فإنه في هذه الحالة:

$$z = \frac{(\bar{x}_1 \pm \bar{x}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}} \sim N(0,1) \ldots\ldots\ldots\ldots. \quad (20-6)$$

أما إذا كان أحد المجتمعين أو كليهما لا يتوزع توزيعاً طبيعياً فإن $(\bar{x}_1 \pm \bar{x}_2)$ لا يتوزع توزيعاً طبيعياً كذلك ، ولكن لقيم $(n_2 \;,\; n_1)$ الكبيرة فإنه طبقاً لنظرية النهاية المركزية السابقة فإن $(\bar{x}_1 \pm \bar{x}_2)$ يتوزع توزيعاً قريباً من التوزيع الطبيعي وبذلك يمكننا استخدام نفس العلاقة (6−20) في حالة العينات الكبيرة.

توزيع المعاينة للتباين Sampling Distribution of The Variance:

إذا كان $S^2 = \frac{\sum(x_1 - \bar{x})^2}{n-1}$ هو تباين عينه عشوائية حجمها (n) مأخوذة من مجتمع متوسطه (μ) وتباينه (σ^2) وعزمه الرابع حول المتوسط هو (μ_4) فإن :

$$\mu_{S^2} = \sigma^2 \qquad and \qquad \sigma^2{}_{S^2} = \frac{(\mu_4 - \sigma^2)}{(n-1)} \;\; \ldots\ldots\ldots\ldots \quad (20-7)$$

وإذا كان المجتمع طبيعي فإن Type equation here. $(\mu_4 = 3\sigma^2)$ وبالتالي فإن :

$$\sigma^2{}_{s^2} = \left(\frac{2}{n-1}\right) \; x \;\; \sigma^2 \;\; \ldots\ldots\ldots.. \quad (20-8).$$

نلاحظ هنا أن σ^2 لا تتوزع توزيعا طبيعيا حتى ولو كان المجتمع طبيعي ، ولكنه يتوزع توزيع قريب من التوزيع الطبيعي وذلك لقيم (n) الكبيرة $(n \geq 100)$. أما إن كان المجتمع الأصلي يخضع للتوزيع الطبيعي فإن المتغير $\frac{(n-1)\,S^2}{\sigma^2}$ يخضع لتوزيع يسمى توزيع مربع كاي χ^2 بعدد درجات حرية يساوي $(n-1)$ أي أن :

$$\frac{(n-1)\,S^2}{S^2} \sim \chi^2\,(n-1) \quad \ldots\ldots\ldots\ldots \quad (20-9)$$

ويعتبر توزيع مربع كاي من التوزيعات الهامة في الإحصاء التطبيقي ودالة كثافته هي :

$$f(y) = y^{\frac{v-1}{2}} \; x \; e^{\frac{-y}{2}} \quad , \quad y > 0 \;\ldots\ldots\ldots\ldots\ldots \quad (20-10)$$

حيث (v) هي عدد درجات الحريه للتوزيع وتعتبر هي المعامل الوحيد له ويتضح من شكل الدالة أنها داله متصلة وتقع بأكملها فوق النصف الموجب لمحور السينات ، منحنى هذه الداله غير متماثل ويعتبر من المنحنيات موجبة الالتواء ويقل التواؤه (وبالتالي يقترب من التماثل) كما زادت درجات الحريه (v). وتكون القيمة المتوقعة لهذا التوزيع هي(v) و تباينه هو $(2v)$ أي بمعنى أن :

$$E(y) = \mu_y = v$$

$$v(y) = \sigma^2 = 2v \quad \ldots\ldots\ldots\ldots \quad (20-11)$$

فإذا كان $({S_1}^2)$ هو تباين عينه عشوائية حجمها (n_1) مسحوبة من مجتمع طبيعي $N(\mu_1, {\sigma^2}_1)$، وكان$({S_1}^2)$ هو تباين عينه عشوائية أخرى حجمها (n_2) ومسحوبة من مجتمع طبيعي آخر $N(\mu_2, {\sigma^2}_2)$ وكانت العينتان مستقلتان فإن المتغير:

$$\frac{{S_1}^2/{\sigma_1}^2}{{S_2}^2/{\sigma_2}^2} \sim \mathrm{F}\,(n_1-1\,, n_2-1\,) \;\ldots\ldots\ldots\ldots\ldots \quad (20-12)$$

حيث أن $\mathrm{F}\,(n_1-1\,, n_2-1\,)$ تسمى بتوزيع (F) بدرجتي الحريه n_1-1 و n_2-1 و دالة الكثافة الإحتماليه للمتغير (y) الذي يخضع لتوزيع (F) بدرجتي الحريه $(v_2\;, v_1\;)$ تعطى بالصورة:

$$f(y) = \frac{y^{\frac{v_1}{2}-1}}{(v1y+v2)^{\frac{v_1\,+\,v_2}{2}}} \quad , \quad y > 0 \;\ldots\ldots\ldots\ldots\ldots \quad (20-13)$$

وكما يتضح من الداله في (13−20) أن المنحنى يقع بالكامل في النصف الموجب لمحور السينات كما في حالة توزيع χ^2 ، وهو أيضاً غير متماثل وموجب الالتواء ولكن يقترب من التماثل كلما زادت درجات الحريه $(v_2\ ,v_1\ )$.

ذكرنا سابقاً أنه إذا كان $\bar{X}$ هو المتوسط الحسابي لعينه حجمها n مأخوذة من مجتمع طبيعي بالمعاملات $(\mu\ ,\sigma^2\ )$ فإن :

$$z = \frac{\sqrt{n}\ \ (\bar{x} -\ \mu)}{\sigma} \quad \sim \quad N\,(0{,}1)$$

هذا إذا كانت (σ) معلومة ، ولكن في حالة ما إذا كانت قيمة (σ) غير معلومة فإننا نستخدم بدلا منها الانحراف المعياري للعينة (S) ولكن في هذه الحالة يصبح المتغير $(\frac{\sqrt{n}\ \ (\bar{x}-\ \mu)}{\sigma})$ يخضع لتوزيع يعرف بتوزيع (t) ستيودنت $(t-student)$ بدرجات حريه $(n-1)$ ، أي أن :

$$t = \frac{\sqrt{n}\ (\bar{x} -\ \mu)}{s} \quad \sim \quad t(n-1) \quad \dots\dots\dots\dots \qquad (20-14)$$

دالة الكثافة لتوزيع (t) بدرجات حريه (v) تعطي بالصورة:

$$f(t) = \left(1 + \frac{t^2}{v}\right)^{\frac{v+1}{2}} \quad \dots\dots\dots\dots \qquad (20-15)$$

وهو توزيع متماثل حول محور (y) وهو يشبه في ذلك المنحنى الطبيعي القياسي $N(0{,}1)$ ولكنه أقل تحدباً من التوزيع الطبيعي القياسي ولكنه يقترب من التوزيع الطبيعي كلما زادت درجات الحريه.

وإذا كان $\bar{X}_2$ و S_1^2 هما المتوسط الحسابي والتباين لعينه حجمها n_1 مأخوذة من مجتمع طبيعي متوسط هو μ_1 وكان $\bar{X}_2$ و S_2^2 هما المتوسط الحسابي والتباين لعينه أخرى حجمها n_2 ومأخوذة من مجتمع طبيعي آخر له المتوسط μ_2 وكانت العينتان مستقلتان فإن المتغير:

$$t = \frac{(\bar{x}_1\ \pm\ \bar{x}_2) -\ (\mu_1\ \pm\ \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \sim \quad t(n_1 +\ n_2\ - 2) \quad \dots\dots\dots \quad (20-16$$

حيث أن :

$S_p{}^2\ =\ \frac{(n_1-1)\ S_1{}^2 + (n_2-1)\ S_2{}^2}{(n_1 + n_2 - 2)}$ يسمى بالتباين المشترك للعينتين The Pooled Variance.

تم

بعون الله تعالى

Arabic References المصادر العربية :

1. عدنان شهاب حمد ومهدي محسن اسماعيل (2001): اساليب المعاينة في ميدان التطبيق , المعهد العربي للتدريب والبحوث الاحصائية , بغداد.
2. عماد غصاب عبابنة وسالم عيسى بدر (2007): مبادئ الاحصاء الوصفي والاستدلالي , الطبعة الاولى , دار المسيرة للنشر والتوزيع والطباعة , عمان .
3. عبدالرحمن محمد ابو عمه وآخرون (1995): مقدمة في المعاينة الاحصائية , دار المريخ للنشر, الرياض .
4. سبيغل, موراي (2001) , الاحصاء والاحتمال, اكاديميا للنشر والطباعة, 420 ص

English References المصادر الانكليزية:

1. Spiegel, R. Murray (1972): Theory and Problems of Statistics, Schaums Outline Series, McGraw-Hill Company.P422.
2. Pearson, E. S. and Kendal, Sir M. (Editors) (1978) " Studies in The History of Statistics and Probability",Vol. I, Charles Griffin , London.
3. Wikipedia, the free encyclopedia Web site: http://en.wikipedia.org/ .

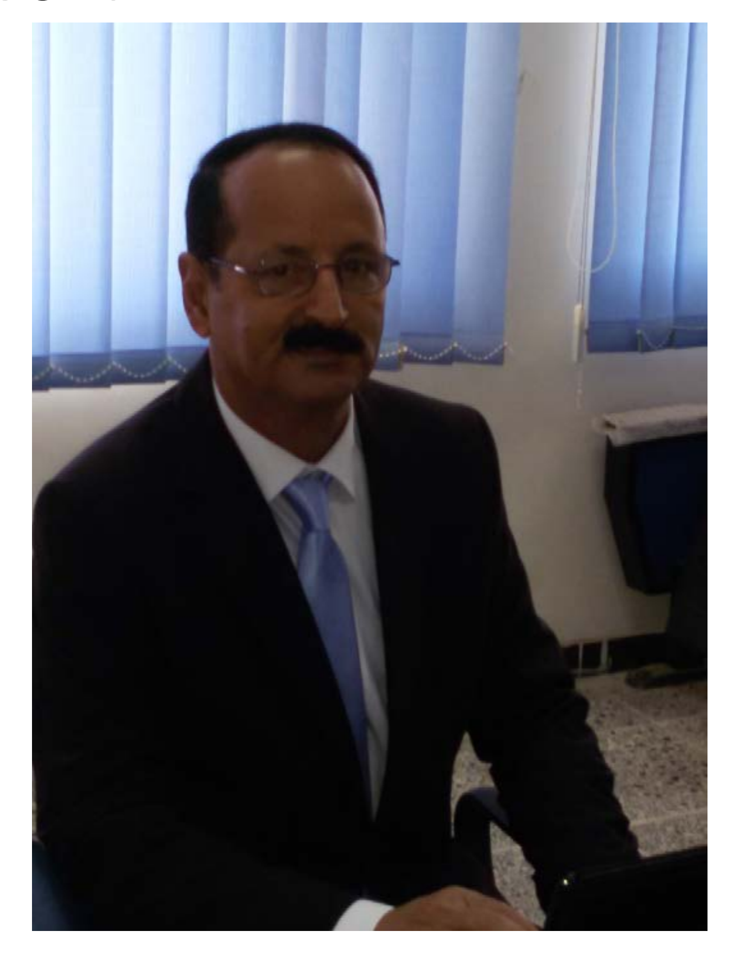
مع تحيات د. سلام حسين الهلالي

salamalhelali@yahoo.com

https://www.facebook.com/salam.alhelali

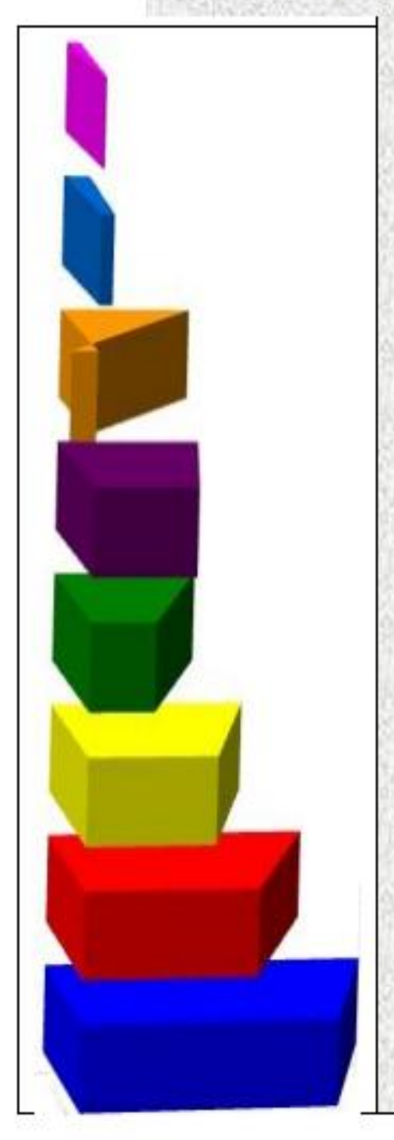
https://www.facebook.com/groups/
/Biothesis

https://www.researchgate.net/profile/
/Salam_Ewaid
07807137614





سيرة المؤلف



- غازي عطية زراك
- مدرس
- التولد: 1956 / محافظة واسط / قضاء الصويرة
- بكالوريوس جيولوجي / جيوفيزياء من قسم علوم الارض / كلية العلوم / جامعة بغداد / عام 1978 .
- ماجستير في جيولوجيا المناجم والاستكشاف المعدني / جامعة لستر / انكلترا / عام 1988.
- عين اول مرة في الشركة العامة للمسح الجيولوجي والتعدين / بغداد / عام 1980 . تدرج في الوظيفة لغاية رئيس جيولوجيين اقدم.
- عمل في برامج الاستكشاف المعدني والجيوفيزيائي وتقييم الترسبات المعدنية وعمليات الاستخلاص المعدني في مناطق مختلفة من العراق لغاية عام 2006.
- انتقل الى جامعة تكريت / كلية العلوم / قسم علوم الارض التطبيقية / نهاية عام 2007، واختص في تدريس مادة جيولوجيا المناجم والاستكشاف المعدني والاستكشاف الجيوفيزيائي فضلا عن تدريس مادة الاحصاء ولحد الان.
- لديه العديد من مؤلفات الكتب العلمية المصدرية والمساعدة في تخصص جيولوجيا المناجم والاستكشاف المعدني.